The Dissertation Committee for Kien Trung Truong
certifies that this is the approved version of the following dissertation:

Transmission Strategies for
Wireless Multiple-Antenna Relay-Assisted Networks

Committee:

Robert W. Heath Jr., Supervisor

Brian L. Evans

Todd E. Humphreys

Scott Nettles

Sujay Sanghavi

Transmission Strategies for Wireless Multiple-Antenna Relay-Assisted Networks

by

Kien Trung Truong, B.S., M.S.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2012

Dedicated to my family.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my family for their unconditional love, support, and sacrifices. They always believe in me and made me the person I am today. I am especially indebted to my wife and my daughter, who endured my long leave-of-absence for most of my Ph.D. program.

I would also like to thank my advisor, Prof. Robert W. Heath Jr., for his constant support. He was kind enough to have me join his research group, and ultimately introduced me to relay communication, the topic of this dissertation. His motivation and technical advice guided the progress in my graduate studies. I owe most of what I learned about research, academic integrity, and intellectual curiosity to him.

Finally, I would like to thank the Vietnam Education Foundation for sponsoring part of my graduate studies. It has been a pleasure researching with my WNCG colleagues, especially the WSIL members. I am also grateful for my Austin friends, who made that city my second home.

Transmission Strategies for
Wireless Multiple-Antenna Relay-Assisted Networks

Kien Trung Truong, Ph.D.
The University of Texas at Austin, 2012

Supervisor: Robert W. Heath Jr.

Global mobile data traffic has more than doubled in the past four years, and will only increase throughout the upcoming years. Modern cellular systems are striving to enable communications at high data rates over wide geographical areas to meet the surge in data demand. This requires advanced technologies to mitigate fundamental effects of wireless communications like path-loss, shadowing, small-scale fading, and interference. Two of such technologies are: i) deploying multiple antennas at the transmitter and receiver, and ii) employing an extra radio, called the relay, to forward messages from the transmitter to the receiver. The advantages of both technologies can be leveraged by using multiple antennas at the relay, transmitter, and receiver. Multiple-antenna relay-assisted communication is emerging as one promising technique for expanding the overall capacity of cellular networks.

Taking full advantage of multiple-antenna relay-assisted cellular systems requires transmission strategies for jointly configuring the transmitters and receivers based on knowledge of the wireless propagation medium. This dissertation proposes such transmission strategies for wireless multiple-antenna relay-assisted systems. Two popular types of relays are considered: i) amplify-and-forward relays (the relays simply apply linear signal processing to their observed signals before retransmitting) and ii) decode-and-forward relays (the relays decode their observed signals and then re-encode before retransmitting). The first part of this dissertation considers the three-node multiple-antenna amplify-and-forward relay channel. Algorithms for adaptively selecting the number of data streams and subsets of transmit antennas at the transmitter and relay to provide reliable transmission at a guaranteed rate are proposed. Expressions for extracting spatial characteristics of the end-to-end multiple-antenna relay channel are derived. The second part of the dissertation presents interference management strategies that are developed specifically for two models of multiple-antenna relay interference channels where a number of relays assist multiple transmitters to communicate with multiple receivers. One model uses amplify-and-forward relays while the other uses decode-and-forward relays. Based on the idea of interference alignment, these strategies aim at maximizing the sum of achievable end-to-end rates. Simulation results show that the proposed transmission strategies with multiple-antenna relays achieve higher capacity and reliability than both those without relays and those with single-antenna relays.

Table of Contents

List of Tables

List of Figures

List of Acronyms

3GPP	The 3rd Generation Partnership Project
AF	Amplify-and-Forward
ASIC	Application-Specific Integrated Circuit
BF	Beamforming
BS	Base Station
CF	Compress-and-Forward
CoMP	Coordinated Multi-Point transmission
CSI	Channel State Information
DAS	Distributed Antenna Systems
DF	Decode-and-Forward
EF	Estimate-and-Forward
FPGA	Field Programmable Gate Array
IEEE	Institute of Electrical and Electronics Engineers
KKT	Karush-Kuhn-Tucker
LTE	Long-Term Evolution
MIMO	Multiple-Input Multiple-Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
MS	Mobile Station
MSE	Mean Squared Error
NNUB	Nearest Neighbor Upper Bound
OFDM	Orthogonal Frequency Division Multiplexing
QAM	Quadrature Amplitude Modulation
QCQP	Quadratically Constrained Quadratic Programming
RC	Relay Channel
RF	Radio Frequency
RS	Relay Station
SDP	Semidefinite Programming
SDR	Semidefinite Relaxation
SF	Selfish beamforming
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio

SR	Sum-Rate
TDMA	Time Division Multiple Access
TH	Two-Hop channel
TL	Total Leakage
VSER	Vector Symbol Error Rate
ZF	Zero Forcing

Chapter 1

Introduction

Mobile devices have become an integral part of our everyday lives due to the high penetration of mobile phones (especially data-intensive smartphones) and emerging applications (such as mobile video) [1, 3]. It is predicted that on average each person on the earth will have a mobile-connected device by the end of 2012 and 1.4 mobile-connected devices in 2016 [2]. The global mobile data traffic has more than doubled annually in the past four years and that rate is expected to keep increasing throughout the upcoming years [3]. The efficiency of cellular networks must be improved to meet the surge of data demand.

Achieving high data rates and good coverage in cellular networks is impeded by fading (small-scale and large-scale) and interference, two fundamental characteristics of the wireless propagation medium [67, 211]. Small-scale fading refers to rapid fluctuations in the strength of radio signals caused by multi-path and Doppler spread effects. Due to reflections from physical objects, radio signals may travel through multiple paths from the transmitter to receiver. The signals on the paths have different time delays, phases and amplitudes. Doppler spread is caused by the relative mobility of the transmitter, receiver and surrounding environment, leading to different shifts in the frequency of radio signals on different paths. Due to small-scale fading, the receiver sees multiple copies of the same signal sent by the transmitter. These copies can add constructively or destructively, resulting in severe fluctuations in the received signal strength. Large-scale fading includes path-loss and shadowing. Path-loss refers to the exponential attenuation of radio signal power with distance. Shadowing refers to the drop in signal power due to large obstructions, such as buildings or trees, obscuring the main path between the transmitter and receiver.

Wireless communication allows multiple users to share common radio frequency resources, making their transmissions interfere with each other. In addition to local thermal noise, each receiver sees more than one signal at the same time, or over the same frequency, making it difficult to decode the desired signal. Cellular systems permit reusing of operating frequency in multiple cells to increase overall network capacity. Recent standards like 3GPP LTE/LTE-Advanced propose the universal frequency reuse, i.e., all cell sites use the same frequency to obtain the most from scarce spectrum resources. This, however, makes cellular systems susceptible to interference.

Multiple-input multiple-output (MIMO) communication and relay communication are two promising techniques to mitigate the effects of fading and interference to support high data rates and reliability. In MIMO communication systems, also known as multiple-antenna systems, the transmitters and receivers are equipped with multiple antennas [61, 62, 199, 201, 202]. In relay communication systems, extra intermediate nodes, also called relays, are deployed to forward messages from the transmitters to receivers [50, 113, 114, 180, 181, 213]. These two techniques can be combined by employing multiple antennas at the transmitters, relays, and receivers, resulting in multiple-antenna relay-assisted systems. In this dissertation, I focus on how to take full advantage of multiple-antenna relay-assisted systems.

This introductory chapter provides background on MIMO communication in Section 1.1 and background on relay communication in Section 1.2. Section 1.3 describes wireless multiple-antenna relay-assisted systems. It also presents the motivations and challenges of my research. Section 1.4 explains the importance of my research in connection with an overview of methods for expanding the overall capacity of cellular networks. This chapter concludes with a thesis statement, a summary of the contributions, notation, and organization of the remainder of this dissertation.

1.1 Wireless Communication with Multiple Antennas

Employing multiple antennas at the transmitters and receivers enables high data rates and reliability in wireless systems. Now that multiple-antenna communication has a relatively mature theoretical foundation (especially for the point-to-point channel without interference), it has been included in several recent standards, including IEEE 802.16e, IEEE 802.11n, and 3GPP LTE/LTE-Advanced [118]. It is expected that the technology will continue to find wide application in future wireless networks.

Multiple antennas can be exploited in many ways in cellular systems including improving the capacity and reliability of the point-to-point transmission between a transmitter and a receiver. With enough spacing between antennas, the fading channel coefficients between different transmit antennas and receive antennas are approximately independent of each other. Thus, the channel matrix between the transmitter and receiver is full rank with a high probability, making it possible to decompose this channel into multiple independent paths. High data rates can be achieved by splitting data into multiple streams and then simultaneously sending the streams over these paths [61, 62, 199, 201, 202]. The number of data streams that can be transmitted simultaneously is referred to as the multiplexing gain. Multiple antennas can also be used to increase transmission reliability by sending multiple copies of a signal through different antennas. There is a high probability that one of these paths is not in deep fade, allowing the receiver to decode this signal successfully. The increase in reliability is referred to as the diversity gain.

Knowledge of wireless channels, also known as channel state information (CSI), is crucial to realizing the benefits of multiple antennas. CSI at transmitters enables the design of strategies to adapt transmitted waveforms to the wireless channel to overcome fading effects to increase received signal strength. For example, in the transmit antenna selection technique, CSI is used to select subsets of transmit antennas corresponding to paths with favorable fading to receive antennas [86, 139, 172]. CSI is also used to design the beamforming vectors (for single-stream transmissions) or the precoding matrices (for multiple-stream transmissions) to adapt the weights of signals while mapping them to transmit antennas [111, 130, 171]. While CSI availability is optional at the transmitters, it is often required at the receivers to estimate wireless channel distortions for reconstructing the transmitted signal. When wireless channels vary slowly enough, it is reasonable to assume that CSI is known instantaneously and perfectly at both the transmitters and receivers. This full CSI assumption is useful in cellular system design since it provides insights into upper-bounds on the performance measures, such as system capacity and reliability.

Interference in cellular systems degrades data rates and increases outages [14, 24, 31, 43], especially when it is not coordinated [8, 75]. Multiple antennas can be used to mitigate interference. Network MIMO is a technique for mitigating interference in cellular networks by forming a large virtual MIMO system from multiple pairs of base stations and mobile stations [136, 231]. This technique is also known in the literature as coordinated multi-point transmission (CoMP) in 3GPP LTE/LTE-Advanced or base station cooperation. This requires base stations exchange control-level signals, and/or transmit data, and/or CSI of both desired and interfering channels, using backhaul communication links [20]. Exploiting the exchanged information, base stations coordinate their transmissions to increase data rates and reduce outages [9, 21, 22, 56, 101, 116, 173, 191, 192, 227, 229, 230]. Much of the prior work on base station cooperation assumed backbone links have infinite capacity and no delay. This allows for simultaneously sending all network data so that interference is precanceled at each receiver. In fact, having base stations communicate through backbone in realtime is a strict and expensive requirement, especially when base stations have to share actual transmitted and/or received symbols to each other. Thus, base station cooperation may not be feasible throughout the entire network. This motivates prior work on reducing the size of cooperation clusters [25, 26, 34, 45, 177, 231]. This also motivates the need of investigating systems that do not support actual transmitted symbol sharing.

Recently, the interference channel has attracted much research interest. It models networks with either no wired connections or low-bandwidth wired connections between the transmitters, thus sharing actual transmitted symbols is infeasible. Each transmitter has data for only one receiver and each receiver is served by only one transmitter. The capacity of the interference channel is not widely known even for the simplest case of two users. Notably, recent work has shown that a physical layer transmission strategy, called interference alignment, is optimal in terms of maximizing the multiplexing gain of the interference channel [27, 64, 135]. In principle, this technique exploits extra time, frequency, or space dimensions in the channel to mitigate interference. For example, space dimensions in the MIMO interference channel, where the transmitters and receivers have multiple antennas, can be used for aligning interference. Specifically, when there are enough antennas, transmitted signals can be constructed so that interference is constrained within only a portion of the receive space at each receiver, leaving the remaining portion for interference-free transmissions [166, 225]. Following the concept of interference alignment, many algorithms have been proposed for cooperatively designing the precoders and decoders in the MIMO interference channel [70, 152, 158, 167, 174, 196, 198]. Although there remain many research challenges, especially practical implementation issues, interference alignment is a promising interference management strategy for cellular networks.

1.2 Wireless Relay Communication

Relay communication is a technique to overcome the effects of fading in wireless networks. In the simplest model, it involves the use of an extra radio, which is wirelessly connected to a transmitter and is called the relay, to forward the message to a receiver, as first studied in [213] and illustrated in Fig. 1.1. Thus, in addition to the (single-hop) direct channel between the transmitter and the receiver, the message travels through the two-hop channel from the transmitter via relay to receiver. Despite extensive research, the capacity of the relay channel is not known in general.

Fig. 1.2 illustrates the benefits of relay communication in extending radio range and combatting shadowing in cellular systems. Two relays RS1 and RS2 assist a base station (BS) to communicate with two mobile stations MS1 and MS2, respectively. Note that for reliable detection and decoding, the desired signal received at a mobile station must have a power value sufficiently above local thermal noise power level. Unfortunately, signals travelling directly from BS to the mobile stations attenuate significantly because of either

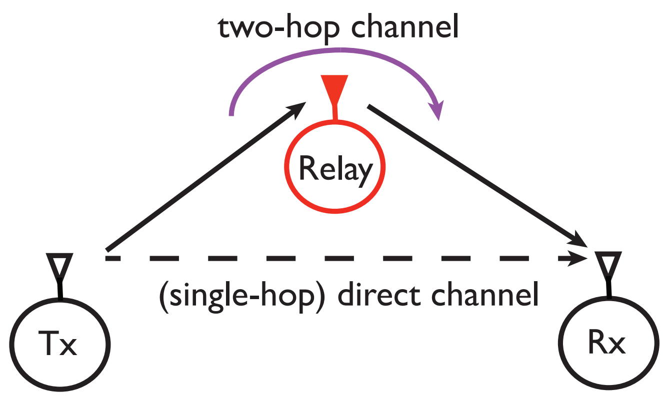

Figure 1.1: The three-node relay channel where a relay assists a transmitter (Tx) to communicate with a receiver (Rx). The dashed line represents the (single-hop) direct channel while the solid line represents the two-hop channel.

high path loss (as MS1 is located outside the boundary of BS's cell area) or shadowing (as there is a building between BS and MS2). Thanks to better geographical locations of RS1 and RS2, signals traveling via the relays experience less power attenuation on BS-relay hop and relay-MS hop, thus improving the received signal strength at MS1 and MS2. Of course, in terms of coverage extension, relays are not needed when mobile stations, e.g., MS3, could receive strong enough signals directly from BS. Thus, relay communication can be viewed as an enabling, or add-on, technology for cellular systems [149].

Analog repeaters, the simplest form of relays with little signal processing capability, have been used widely in 2G/3G cellular systems for coverage extension. These relays simply apply a fixed amplifying gain to both the observed signal and local thermal noise before forwarding to mobile stations [55]. This operation increases noise power and suffers from the danger of instability

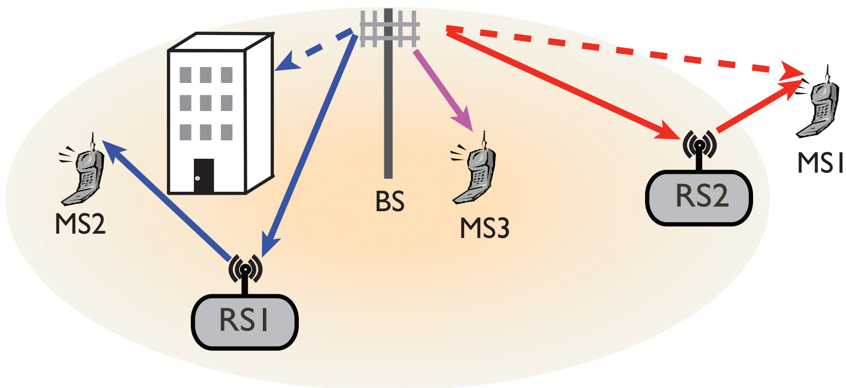

Figure 1.2: Relays (RS1 and RS2) are used for extending communication range and combatting shadowing to aid a base station (BS) to provide services to two mobile stations (MS1 and MS2).

due to the relays' fixed-gain amplifier [148]. Smart relaying has yet to see wide deployment in current cellular systems. Upcoming cellular standards like 3GPP LTE-Advanced and IEEE 802.16m, however, are considering relays with more signal processing capabilities as a viable solution to coverage extension and capacity enhancement [5, 98]. This motivates the need for research on finding and realizing the full benefits of smart relays.

Many different relay transmission strategies have been developed. Relays can be either full-duplex or half-duplex. Full-duplex relays can transmit and receive at the same time while half-duplex relays cannot. Because full-duplex relays are difficult to implement, half-duplex relays are of more interest in both academic and industrial research [19, 159]. Relays are also classified by how they process the received signal. Popular relay types are decode-and-forward (DF–the relay decodes the signal then re-encodes before retransmitting) and amplify-and-forward (AF–the relay simply applies linear transformation to the signal before forwarding).

DF relays have higher computational complexity due to the requirement of decoding their received signals. They are only helpful if they can successfully decode the signals. Transparent to the modulation and coding of the signals, AF relays can be used flexibly in networks comprising of many nodes of different complexity or standards [19]. Also known as non-regenerative relays, AF relays provide better diversity performance, and in some cases higher end-to-end throughput (e.g., when the transmitter-relay channel is weak), than DF relays in single-antenna half-duplex relay systems [29, 114]. AF relays may be attractive in practice thanks to their lower complexity and faster signal processing. DF relays, however, still have a place in upcoming and future cellular systems. When the channels between base stations and relays are strong, DF relays can reliably detect the received signals and then provide higher end-to-end throughput than AF relays. One key limitation of AF relays is the noise propagation effect, i.e., local noise at relays is amplified and forwarded to receivers. Moreover, DF relays act as conventional users while receiving data from base stations. They also act as conventional base stations while transmitting to users. Thus, DF relays can be integrated into current cellular systems with less required changes than AF relays. Given their pros and cons, this dissertation considers both AF and DF relays.

Much prior work on relays, including IEEE 802.16j standard [97, 155], considered single-antenna relay-aided systems because relays were mainly envisioned to provide coverage extension. Investigating potential gains of relays in cellular networks has begun. In particular, the gains provided by IEEE 802.16j relays in cellular networks over the direct transmission have been numerically analyzed using system simulators [18, 93, 148, 159, 179, 214, 226], idealized terrain [54], ray-tracing software applied to urban areas [99, 178], and experiments with an LTE-Advanced testbed [217]. The general conclusion is that single-antenna relay communication is useful for coverage extension, but it provides marginal capacity gains. The main reason is that the capacity gains are limited by interference while many prior strategies for single-antenna relay systems were designed under interference-free assumptions.

1.3 Multiple-Antenna Relay-Assisted Systems

Relays have yet to be a huge success in cellular systems partly because much of their potential has not been fully exploited. One advantage of relays is that they are readily combined with other technologies [114]. For example, MIMO concepts can be incorporated into relay transmissions by deploying multiple antennas at relays, along with transmitters and receivers [133]. Commercial wireless systems, such as 3GPP LTE-Advanced and IEEE 802.16m, are considering multiple-antenna relay communication a viable solution to provide better coverage for high data rate services [4–6, 98, 127, 223].

Multiple-antenna relay-assisted systems provide the high capacity and reliability capability of MIMO communication with the coverage extension capability of relay transmissions [58, 117, 234]. This is useful to provide reliable services to cell-edge users, those located near the cell boundary. As discussed in Section 1.1, there have been many results for realizing the benefits of using multiple antennas in the MIMO point-to-point channel. Nevertheless, it is challenging to extend these results to the MIMO relay channel. One reason is that the presence of the relay increases the number of constituent point-to-point MIMO channels between a transmitter to its associated receiver. Specifically, the MIMO relay channel consists of the following three MIMO point-to-point channels: i) the transmitter-receiver channel, ii) the transmitter-relay channel, and iii) the relay-receiver channel. This makes it difficult to extract end-to-end spatial characteristics of the MIMO relay channel that are crucial to the design of its advanced transmission strategies. The first part of this dissertation focuses on addressing this problem.

Multiple antennas may be used to mitigate the effects of interference in relay-assisted cellular systems. Much prior work on MIMO relay communication, however, neglects interference and hence focuses on the basic three-node MIMO relay channel [42, 78–80, 96, 125, 126, 141, 169, 200, 215, 222]. In addition to the network topology change, the introduction of relays creates more sources of interference. Thus, it is challenging to extend single-hop interference management strategies to those for the MIMO relay interference channel. In a trivial approach, the MIMO relay interference channel can be treated like a cascade of two independent MIMO single-hop interference channels, i.e., the transmitter-relay channel and relay-receiver channel. Obtaining the most from the MIMO relay interference channel, however, requires the joint configuration of the transmitters, relays, and receivers. This motivates the need for interference management strategies that are designed specifically for relay-assisted cellular networks. The second part of this dissertation proposes interference management strategies for the relay interference channel. These strategies are able to take into account special features of relay communication, such as multi-hop transmission and relay signal processing operation, to maximize the sum of achievable end-to-end rates.

1.4 Overview of Cellular Network Capacity Expansion

This dissertation proposes transmission strategies for obtaining high data rates and reliability of wireless multiple-antenna relay-assisted systems. The importance of the contributions can be explained further by looking at the big picture of how current macro-cell-based cellular networks evolve to meet the surge of mobile data traffic. Specifically, the following four possible technological approaches to doing this have been identified [17]:

- *Offload to other radio technologies* The principle is to offload data traffic to other-technology radio networks covering the same area. For example, WiFi offloading is attractive thanks to the wide availability of WiFi (or IEEE 802.11) access points and WiFi-enabled smartphones. Nevertheless, this approach raises technical challenges such as seamless handover and security [138]. Tools for attacking IEEE 802.11 networks are more accessible than those for attacking cellular networks.

- *Add spectrum to cellular networks* For example, multiple LTE carriers may be aggregated on the physical layer to provide a higher bandwidth (up to 100MHz) for an LTE-Advanced user [153]. Unfortunately, spectrum is a scarce commodity, especially for the most attractive frequency bands, thus licenses have become increasingly expensive.

- *Advanced radio link transmission and reception techniques* Many increasingly complex communication techniques have been proposed for successfully sending more bits per second for a given bandwidth over an isolated point-to-point link. Unfortunately, radio link level improvements like error-correction coding, orthogonal frequency division multiplexing (OFDM), iterative receivers, and multiple antennas are reaching their theoretical performance limits [76, 216, 233].

- *Increasing cell density* This approach aims at improving the area spectral efficiency of cellular networks, which is defined as the average sum of data rates per unit bandwidth and per unit area supported by a cell site [11]. A conventional method for doing this is cell splitting, i.e., deploying multiple macro-cells to cover an area where there used to be only one. The macro-cell deployment, however, has become prohibitively expensive in many dense urban areas, including important markets like New York and San Francisco [1, 108, 134]. Network level improvements like base station cooperation and interference alignment are subject to this limitation.

Carriers need to keep cost per bit low while upgrading their cellular networks. Unfortunately, these approaches are reaching practical limits in many dense urban areas and do not provide significant capacity enhancement [13, 128]. In addition, without changing the homogeneous topology of current macro-cell-based cellular networks, such approaches may not work well due to low signal-to-interference-plus-noise ratio (SINR) conditions. Consequently, carriers are revisiting the conventional cellular system topology and are considering a new paradigm obtained by installing low-power and less-expensive nodes into current networks. This new infrastructure is known as heterogeneous networks [49, 108]. In principle, this brings the infrastructure closer to users by shortening transmission distance. Thus, it improves radio link quality and spectrum reuse efficiency, leading to higher area spectral efficiency. Heterogeneous networks are expected as one of the major network capacity enhancement techniques for upcoming cellular standards like 3GPP LTE/LTE-Advanced.

Different types of low-power nodes can be used, including pico-cells, femto-cells, fixed relays, and distributed antennas. Each type of low-power node has its own capabilities, constraints, and operational functionalities. Notably, they are envisaged to coexist in the same geographical area and share the same spectrum, forming multi-tier cellular network roll-outs [128]. Thus, they should be considered complimentary, rather than competing, infrastructure components of upcoming cellular standards. Moreover, depending on the corresponding wireless access technologies, each type of low-power node brings into play different technical challenges and hence attracts a lot of research interests on its own. For example, Section 1.3 presents the challenges posed by the introduction of relays into cellular networks.

In this dissertation, I focus only on one type of low-power node, which is fixed relays. As part of the infrastructure, fixed relays are connected wirelessly to base stations. Compared to the cell splitting approach, the use of relays eliminates the costs of deploying and maintaining backhaul transmission lines connecting new cell sites and the wired backhaul network. While a conventional base station requires a transmit power of 46 dBm to serve a cell of diameter 2-5 km, a relay requires a transmit power of 30 dBm to cover a region with a diameter 200-500 m [128, 148]. Technically, this low transmit power requirement allows for economical design of relays in relative comparison with conventional base stations. Furthermore, serving smaller cell areas, relays can be installed on lower masts than base stations, thus relatively reducing operating expenses like tower leasing and maintenance costs.

My contributions in this dissertation aim at realizing the full potential of fixed relays as a viable solution to increase area spectral efficiency of cellular networks. They will be enumerated in Section 1.6 and presented in detail in the following chapters. In the long term, cellular networks may evolve to heterogenous networks with a mixture of all types of low-power nodes to meet the surge of mobile data traffic. These contributions provide insights and foundations for future work on realizing the full potential of such mixed heterogeneous networks.

1.5 Thesis Statement

Transmission strategies for jointly configuring the transmitters and relays based on knowledge of channel state information improve data rates and reduce outages in multiple-antenna relay-aided communication systems.

1.6 Contributions

In this dissertation, I develop transmission strategies for jointly configuring the transmitters and relays in several configurations of wireless multiple-antenna relay-aided communication systems. These may be applicable for upcoming and future wireless standards like the ones studied by the IEEE 802.16 and 3GPP LTE-Advanced standard bodies. The contributions in this dissertation can be summarized as follows:

(1) I propose adaptive transmit antenna selection algorithms for the MIMO AF relay channel where an AF relay aids the transmission between a transmitter-receiver pair [205, 206]. The design objective is to improve the reliability of fixed-rate transmissions. This is relevant for providing services at a guaranteed rate to users located near or even beyond the boundary of a base station's cell area.

- I use tools from matrix analysis to derive expressions that extract end-to-end spatial characteristics of the MIMO AF relay channel. Acting as the regular condition number of the MIMO point-to-point channel [84], these expressions are named the cascade condition number when the direct link is ignored and the relay condition number when the direct link is considered.

- I develop and analyze algorithms for selecting the number of data streams and subsets of transmit antennas at the transmitter and relay to improve link reliability.
- I propose and analyze algorithms for switching between spatial multiplexing and selection diversity in the MIMO AF relay channel.

(2) I develop three cooperative algorithms for jointly configuring the transmitters and relays in the MIMO AF relay interference channel [207, 210]. The relay interference channel models a network where a stage of relays assist multiple transmitters to communicate with their receivers using shared radio resources.

- I develop and analyze an algorithm for jointly designing the transmitters and relays so that interference and enhanced noise from the relays are aligned and canceled. It is inspired by the interference alignment algorithms for the MIMO single-hop interference channel in [69, 156].
- I propose and analyze two algorithms for jointly designing the transmitters and relays for solving the end-to-end sum-rate maximization problems with either equality or inequality power constraints. These algorithms are guaranteed to converge to the stationary points

of the corresponding design problems. They are generalized versions of the algorithm for the MIMO single-hop interference channel in [177].

- Simulation results show that AF relays enhance the feasibility of interference alignment in the MIMO relay interference channel, leading to higher end-to-end multiplexing gains than both DF relays and direct transmission.

(3) I propose a three-phase algorithm for jointly designing the transmit precoders at the transmitters and relays in the MIMO DF relay interference broadcast channel [209]. In this model, multiple transmitters communicate with their associated receivers with the aid of a stage of relays. Each relay simultaneously forwards data from a single transmitter to multiple receivers. Each transmitter may require the aid of multiple relays at the same time. Each receiver is served by only one transmitter via a single relay. In the simplest form, this model is treated like a cascade of two MIMO single-hop interference broadcast channels.

- I notice that in addition to interference, a mismatch between the achievable rates on two hops could lead to low end-to-end sum-rates in DF relay networks [208, 209]. By definition, a two-hop rate mismatch occurs if there coexist two-hop links with a dominant first hop and those with a dominant second hop.

- I dvelop and analyze a three-phase algorithm that simultaneously mitigates interference and matches the rates on two hops to find high-quality suboptimal solutions in terms of end-to-end sum-rate maximization. The proposed algorithm allows for distributed implementation with low overhead and has fast convergence [209].

1.7 Notation

I use the following notation throughout this dissertation. Normal letters (e.g., a) are used for scalars. Bold lowercase and uppercase letters (e.g., $\mathbf{h}$ and $\mathbf{H}$) represent column vectors and matrices, respectively. $\mathbb{R}$ is the set of real numbers while $\mathbb{C}$ is the set of complex numbers. $\mathbf{I}_N$ and $\mathbf{0}_N$ are the identity matrix and all-zero matrices of size $N \times N$. The $(m, n)^{\text{th}}$ element in a matrix $\mathbf{A}$ is denoted by $[\mathbf{A}]_{mn}$. For a matrix $\mathbf{A}$, $\mathbf{A}^T$ is the transpose matrix, $\|\mathbf{A}\|_F^2$ the Frobenious norm, $\mathbf{A}^*$ the conjugate transpose, $\mathrm{tr}(\mathbf{A})$ the trace, and $\mathbf{A}^\dagger$ is the pseudo-inverse. $\|\mathbf{h}\|_{\mathrm{F}}$ stands for the Frobenius norm of $\mathbf{h}$, while $\|\mathbf{H}\|_{\mathrm{F}}$ is that of $\mathbf{H}$. $\mathrm{vec}(\mathbf{A})$ denotes the vec operator to transform $\mathbf{A}$ into $\mathbf{a}$ while $\mathrm{vec}^{-1}(\mathbf{a})$ denotes the inverse operator. The complex zero-mean Gaussian distribution is denoted by $\mathcal{CN}(\mathbf{0}, \mathbf{Y})$, where $\mathbf{Y}$ is the covariance matrix. $\otimes$ is the Kronecker product. $\mathbb{E}[\cdot]$ is the statistical expectation operator. $()^{(n)}$ denotes iteration index. $()_{\mathrm{T}}$ is used for transmitters' parameters, $()_{\mathrm{R}}$ for receivers', and $()_{\mathrm{X}}$ for relays'. $\triangleq$ is an equation-by-definition. $\mathbb{E}[\cdot]$ is the statistical expectation operator. Further notation is introduced in the chapters, as the need arises.

1.8 Organization of Dissertation

The reminder of this dissertation is organized as follows. In Chapter 2, I propose several algorithms for adaptively selecting the number of data streams and subsets of transmit antennas at the transmitter and relay in the three-node multiple-antenna AF relay systems. In Chapter 3, I present cooperative interference management strategies for jointly designing linear precoders at the transmitters and relays in the multiple-antenna AF relay interference channel where each relay is dedicated to assisting a single transmitter-receiver pair. In Chapter 4, I address the problem of jointly configuring the linear precoders at the transmitters and relays in the MIMO DF relay interference broadcast channel where each relay may forward data from a transmitter to multiple receivers. In Chapter 5, I conclude this dissertation with a summary of the results and suggestions for future work.

Chapter 2

Multimode Antenna Selection for MIMO Amplify-and-Forward Relay Systems

This chapter proposes algorithms for adaptive transmit antenna selection for multiple-antenna AF relay systems. Section 2.1 presents the motivations, reviews prior work, and introduces contributions. Section 2.2 describes the system model. Section 2.3 provides general vector symbol error rate (VSER) analysis and relay filter design for both the two-hop channel and relay channel. While Section 2.4 presents the detailed analysis and the proposed algorithms for the two-hop channel, Section 2.5 presents those for the relay channel. Section 2.6 provides Monte Carlo simulations to evaluate the VSER achieved by the proposed algorithms.

2.1 Introduction

Multiple-input multiple-output relay systems provide the high capacity of MIMO communication with the coverage extension capability of relay transmission [58, 117, 234]. Upcoming commercial wireless systems, such as 3GPP LTE-Advanced and IEEE 802.16m, are considering MIMO relay communication a viable solution to providing better coverage for high data rate services [5, 98]. Taking full advantage of MIMO relay systems requires jointly configuring the transmitter and relay adaptively to current channel conditions. This chapter presents adaptive algorithms for selecting the number of data streams and transmit antenna subsets at the transmitter and relay in half-duplex MIMO AF relay systems to improve the reliability of fixed-rate transmissions.

Half-duplex MIMO AF relays have been studied in [78, 109, 141, 157, 200]. The authors of [78, 141, 200] considered spatial multiplexing relay systems where the transmitter simultaneously sends independent data streams with the same power, i.e., the transmit covariance matrix is a scaled identity matrix. Linear filters at the relays were designed either to maximize the end-to-end mutual information [141, 200], or to minimize the mean squared error (MSE) of vector symbol detection [78]. Single-stream transmission strategies that could achieve the full diversity order of the MIMO relay channel were proposed in [109] and [157]. Note that point-to-point spatial multiplexing does not work well at the cell edge [14]. While the point-to-point link may have more diversity thanks to MIMO, it will not be able to get high data rates from spatial multiplexing. A MIMO relay extends the multiplexing capability to users located near the cell edge.

Assuming a fixed number of data streams, the designs in [78, 109, 141, 157, 200], like much prior work on MIMO AF relays, worked well only in certain channel conditions. This motivates the development of relaying protocols that can adapt to channel conditions. There have been many adaptive algorithms for the point-to-point MIMO channel [32, 83, 84]. Their extension to the MIMO relay channel, however, is not straightforward since the relay introduces an additional dimension in mode adaptation. Furthermore, the relay channel includes contributions from three point-to-point matrix channels. Thus, the simple condition number analysis of the point-to-point MIMO channel, as in [83], does not apply. Many existing adaptive relay protocols are applicable only to single-antenna relays [23, 114, 212]. To the best of my knowledge, the only adaptive algorithm for MIMO AF relay systems so far was proposed in [154]. Similar to my proposed dual algorithms, it considered only two modes with the same total number of transmitted bits. One mode is full spatial multiplexing (i.e., all antennas are used) and another is full selection diversity (i.e., single-stream transmission using the best antenna at the transmitter and relay). The principle is based on a sufficient condition for full spatial multiplexing to have a lower VSER than full selection diversity. The main difference is that the algorithm in [154] considered nonlinear maximum-likelihood (ML) receive filters at the receiver while mine considers linear zero-forcing (ZF) receive filters, which are practically attractive due to their low complexity.

The benefits of MIMO communication are obtained at a price of complexity, size, and cost that scales with the number of active antennas. Transmission of each data stream requires a radio frequency (RF) chain, which is relatively more expensive than antenna elements and digital signal processing components. Moreover, in relay-aided cellular networks, due to space and cost constraints, it is likely that mobile stations have fewer antennas than relay stations and base stations. This means that the number of data streams to be processed at either relay stations or base stations will be less than the number of available antennas. Antenna selection is a technique that enables the use of fewer RF chains than antenna elements, where only the best subset of antennas are selected based on the channel states. Antenna selection for the point-to-point MIMO channel has been studied extensively, see [139, 172], and references therein. In many scenarios, antenna selection leads to significant savings while incurring a (usually small) performance loss compared to the full-complexity systems (with the same number of antennas but each has its own RF chain). In addition, in the feedback context, antenna selection reduces overhead since only the channel information related to the selected antennas needs to be sent back.

In this chapter, I develop algorithms for adaptive transmit antenna subset selection at the transmitter and relay in half-duplex MIMO AF relay systems with linear ZF receive filters. An antenna selection mode of operation, or a mode for short, is defined by the number of selected transmit antennas at the transmitter (which is equal to the number of data streams), the stream-to-antenna mapping at the transmitter, the number of selected transmit antennas at the relay, and the stream-to-antenna mapping at the relay. Having the perfect instantaneous information of all the constituent channels [141, 200], the receiver selects dynamically the mode that minimizes approximations of the VSER for a given overall data rate. The selected mode index is sent back to the transmitter and relay via low-rate feedback channels. The transmitter and relay configure their parameters according to the feedback. My work is thus a natural, but not straightforward, generalization of the results in [83, 154]. In particular, precoding and antenna selection for the MIMO relay channel are still not developed as extensively as those for the point-to-point case. Part of the novelty of this chapter is the problem formulation. Another novelty is the mathematical simplifications (generalizations of the concept of condition number), which are not straightforward.

My approach of adaptation aims to minimize the VSER at a fixed overall data rate. This is motivated by the need for reliable transmissions at a guaranteed rate. For example, this approach is relevant for providing services like voice and audio or for providing the minimum levels of service to users located near the boundary of the cells. The system design objective in this case is often to maximize the transmission range subject to a threshold on error-rate performance at a fixed rate [117, 234]. The design problems with the same objective have been investigated for both point-to-point MIMO systems [107, 111, 171, 176] and MIMO relay systems [154]. This chapter considers the following two configurations: i) the two-hop channel (when the direct link is too weak and hence can be neglected at the receiver) and ii) the relay channel (when the direct link is good and can be utilized by the receiver) [50, 213]. For each configuration, I develop adaptive antenna selection algorithms that aim at selecting the mode with the lowest VSER. Because the closed-form expression for the VSER is not available, it is challenging to find the optimal solution. Fortunately, since mode selection criteria require only the relative comparison of VSER values, it is possible to develop suboptimal algorithms based on VSER approximations that still select good modes. The development of the algorithms takes into account computational complexity. Since the exact values for computational complexity depend on specific implementation targets (general processor, digital signal processor, FPGA, or ASIC), I provide a qualitative comparison of the proposed algorithms.

First, I develop dualmode algorithms for systems that support only two special modes: full spatial multiplexing and full selection diversity. Since the latter mode achieves the full diversity gain of the corresponding channel in case of Rayleigh fading channel [157], the optimal dualmode algorithm, if any, should do too. The simulations show that my dualmode algorithms obtain the full diversity order. This means, although suboptimal, to a certain degree, my algorithms can correctly select the better mode. The simulations also show that two out of the three proposed dualmode algorithms outperform the limited feedback Grassmannian beamforming design in [109]. The other algorithm is based on my proposed concepts of the effective condition numbers of the two-hop channel and relay channel. As fuctions of the singular values of constituent channels, these concepts provide intuitive quantifications of the end-to-end channel quality.

Second, I develop adaptive algorithms for systems that support multimode transmission. Note that full spatial multiplexing requires implicitly that the number of RF chains be equal to the number of transmit antennas, somehow blurring the advantage of antenna selection. Multimode systems, however, allow the use of any number of RF chains less than the number of antennas. The link-level simulations show that the proposed multimode algorithms achieve the full diversity gain and provide considerable array gains over the dualmode algorithms and existing strategies. This is because multimode transmission supports in-between modes to provide a finer control of transmission parameters at the transmitter and relay. It has more choices of spatial multiplexing constellations, thus leading to better adaptation of transmitted signals to the channel conditions. Furthermore, the multi-cell simulations, which adopt the setting in [159], show that the multimode algorithms work well in an interference-limited cellular network and improve significantly the VSER performance of downlink transmission to cell-edge users.

For notation convenience, the superscripts $(\cdot)^{(m)}$ and $(\cdot)^{(s)}$ are used to indicate full spatial multiplexing and full selection diversity.

2.2 System Model

In this chapter, I consider a half-duplex AF MIMO relay system where the transmitter wishes to send data to the receiver with the aid of the relay, as illustrated in Fig. 2.1. I assume that the total number of bits carried in a symbol vector is constant over channel realizations, i.e., fixed-rate transmissions. Since half-duplex relays cannot transmit and receive at the same time by assumption, relay-aided transmission needs two stages. In the first stage, the transmitter broadcasts data to both the relay and receiver. In the second

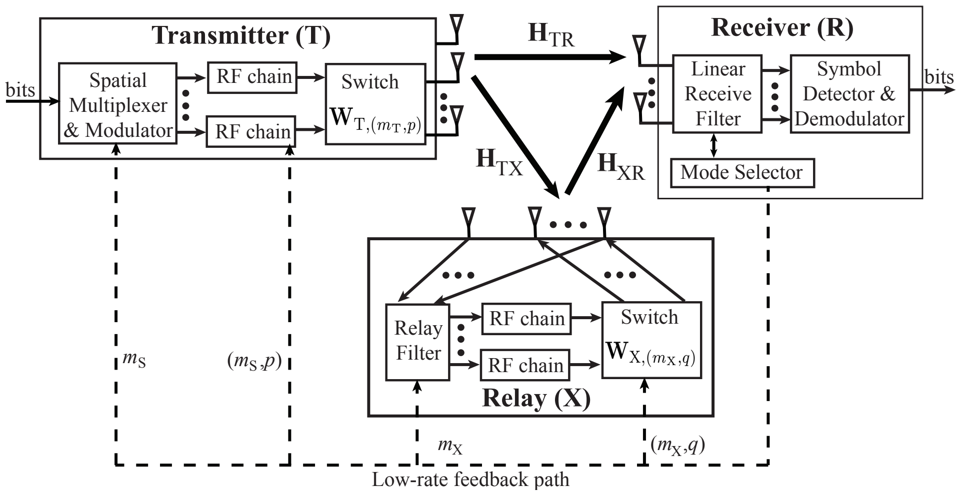

Figure 2.1: Spatial multiplexing amplify-and-forward relay systems with feedback. In the first stage, the transmitter broadcasts the message to the relay and receiver. In the second stage, the relay forwards its observed signal to the receiver, while the transmitter is silent. Having full channel state information, the receiver determines the best mode. The transmitter and relay configure their transmission parameters based on feedback from the receiver.

stage, without decoding its observed signal, the relay applies a linear filter and forwards to the receiver. By assumption, the transmitter either remain silent (e.g., to increase battery life) or receive data from another terminal in the network [142]. In the protocol, the transmitter and relay transmit on orthogonal channels in time or in frequency [109, 114, 141, 200]. In other protocols like non-orthogonal AF relaying, the transmitter may send an extra message to increase spectral efficiency in high signal-to-noise ratio (SNR) regimes [142]. The gain, however, is not realized in low SNR regimes [16]. Such protocols are out of the scope of this chapter.

I assume the transmitter, receiver, and relay have N_T, N_R, and N_X

antennas, respectively. I consider slowly-varying, frequency-flat, block-fading channels. Let $\mathbf{H}_{\mathrm{TR}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ denote the transmitter-receiver matrix channel, $\mathbf{H}_{\mathrm{TX}} \in \mathbb{C}^{N_{\mathrm{X}} \times N_{\mathrm{T}}}$ the transmitter-relay channel, and $\mathbf{H}_{\mathrm{XR}} \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{X}}}$ the relay-receiver channel. I assume Gaussian signaling for tractable analysis although it might not be optimal for MIMO AF relay systems. The received signals are corrupted by additive, circularly symmetric complex white Gaussian noise with zero mean and variance σ^2. Let $\mathbf{n}_{\mathrm{R1}}$ and $\mathbf{n}_{\mathrm{R2}}$ be the noise vectors at the receiver in two stages. Let $\mathbf{n}_{\mathrm{X}}$ be the noise at the relay. Note that $\mathbf{n}_{\mathrm{R1}}, \mathbf{n}_{\mathrm{R2}} \sim \mathcal{CN}(\mathbf{0}_{N_{\mathrm{R}}}, \sigma^2 \mathbf{I}_{N_{\mathrm{R}}})$ and $\mathbf{n}_{\mathrm{X}} \sim \mathcal{CN}(\mathbf{0}_{N_{\mathrm{X}}}, \sigma^2 \mathbf{I}_{N_{\mathrm{X}}})$. I assume that uncoordinated interferers, if any, use independent Gaussian codebooks for their transmissions. While decoding the desired signal, the receiver treats interference from the uncoordinated interferers as independent extra sources of additive Gaussian noise.

This chapter focuses on adaptive transmit antenna subset selection algorithms [139, 172]. Joint antenna selection on both the transmitter and receiver, which further reduces the number of receive RF chains at a price of higher complexity, is left for future work. The receiver and relay have N_{R} and N_{X} receive RF chains. The transmitter and relay have M_{T} and M_{X} transmit RF chains, where $M_{\mathrm{T}} \leq N_{\mathrm{T}}$ and $M_{\mathrm{X}} \leq N_{\mathrm{X}}$. According to antenna selection, m_{T} out of the M_{T} transmit RF chains at the transmitter and m_{X} out of the M_{X} transmit RF chains at the relay are active.

The transmitter consists of a spatial demultiplexer that splits R bits to be transmitted into m_{T} different bit streams (R is assumed to be divisible by m_{T}). The bit streams are modulated independently using the same constellation of size $2^{R/m_{\mathrm{T}}}$ to form a symbol vector $\mathbf{x}_{\mathrm{T}} \in \mathbb{C}^{m_{\mathrm{T}} \times 1}$ such that $\mathbb{E}[\mathbf{x}_{\mathrm{T}}\mathbf{x}_{\mathrm{T}}^*] = (E_{\mathrm{T}}/m_{\mathrm{T}})\mathbf{I}_{m_{\mathrm{T}}}$, where E_{T} is the average transmit energy at the transmitter [61]. Let $\mathcal{M}_{\mathrm{T}}$ denote the set of supported numbers of streams. After going through the transmit RF chains, the m_{T} streams are mapped to a subset of N_{T} transmit antennas. For each m_{T}, the $\binom{N_{\mathrm{T}}}{m_{\mathrm{T}}}$ possible stream-to-antenna mapping matrices formed by m_{T} columns of $\mathbf{I}_{N_{\mathrm{T}}}$ are indexed and denoted by $\mathbf{W}_{\mathrm{T},(m_{\mathrm{T}},p)}$, where $p = 1, 2, \cdots, \binom{N_{\mathrm{T}}}{m_{\mathrm{T}}}$.

This chapter assumes symbol timing errors and frequency offsets are negligible and channel estimation is perfect on the receive side. The effective channels corresponding to the selected antennas at the transmitter are defined as as $\mathcal{H}_{\mathrm{TR},(m_{\mathrm{T}},p)} \triangleq \mathbf{H}_{\mathrm{TR}}\mathbf{W}_{\mathrm{T},(m_{\mathrm{T}},p)} \in \mathbb{C}^{N_{\mathrm{R}} \times m_{\mathrm{T}}}$ and $\mathcal{H}_{\mathrm{TX},(m_{\mathrm{T}},p)} \triangleq \mathbf{H}_{\mathrm{TX}}\mathbf{W}_{\mathrm{T},(m_{\mathrm{T}},p)} \in \mathbb{C}^{N_{\mathrm{X}} \times m_{\mathrm{T}}}$. For notational convenience, the subscripts (m_{T}, p) are suppressed, thus only $\mathcal{H}_{\mathrm{TR}}$ and $\mathcal{H}_{\mathrm{TX}}$ are used. The received signals in the first stage are

$$\mathbf{y}_{\mathrm{R1}} = \mathcal{H}_{\mathrm{TR}}\mathbf{x}_{\mathrm{T}} + \mathbf{n}_{\mathrm{R1}}, \tag{2.1}$$

$$\mathbf{y}_{\mathrm{X}} = \mathcal{H}_{\mathrm{TX}}\mathbf{x}_{\mathrm{T}} + \mathbf{n}_{\mathrm{X}}. \tag{2.2}$$

In the second stage, the relay applies a linear filter $\mathcal{F} \in \mathbb{C}^{m_{\mathrm{X}} \times N_{\mathrm{X}}}$ to $\mathbf{y}_{\mathrm{X}}$ before retransmitting. The design of $\mathcal{F}$ depends on the CSI availability at the relay. In addition to knowledge of $\mathcal{H}_{\mathrm{TX}}$ via channel estimation, the relay is assumed to know $\mathcal{H}_{\mathrm{XR}}$ via feedback, but not the original channel $\mathbf{H}_{\mathrm{XR}}$, thus reducing feedback overhead. For the relay channel, although knowledge of the direct channel may help design a better relay filter for VSER minimization, such a design is unknown and not easy to derive. Even if available, it may lead to intractable analysis. For simplicity and overhead reduction, the following assumptions are made on the availability of channel state information. First, the relay has no information on the direct channel. Second, the receiver has perfect information of all original channels $\mathbf{H}_{\mathrm{TX}}$, $\mathbf{H}_{\mathrm{TR}}$, and $\mathbf{H}_{\mathrm{XR}}$. Finally, the transmitter has no channel state information. The design details of $\mathcal{F}$ are provided in Subsection 2.3.2.

The streams are mapped to a subset of m_{X} antennas via a stream-to-antenna mapping matrix, which is formed by m_{X} columns of $\mathbf{I}_{N_{\mathrm{X}}}$ and denoted by $\mathbf{W}_{\mathrm{X},(m_{\mathrm{X}},q)}$ for $q = 1, 2, \cdots, \binom{N_{\mathrm{X}}}{m_{\mathrm{X}}}$. The effective relay-receiver matrix channel is defined as $\mathcal{H}_{\mathrm{XR},(m_{\mathrm{X}},q)} \triangleq \mathbf{H}_{\mathrm{XR}}\mathbf{W}_{\mathrm{X},(m_{\mathrm{X}},q)} \in \mathbb{C}^{N_{\mathrm{R}} \times m_{\mathrm{X}}}$. Let E_{X} be the average transmit energy at the relay. The power constraint at the relay is

$$\mathrm{tr}\left\{\mathcal{F}\Big(\frac{E_{\mathrm{T}}}{m_{\mathrm{T}}}\mathcal{H}_{\mathrm{TX}}\mathcal{H}_{\mathrm{TX}}^{*} + \sigma^2\mathbf{I}_{N_{\mathrm{X}}}\Big)\mathcal{F}^{*}\right\} \leq E_{\mathrm{X}}. \tag{2.3}$$

In this stage, after sampling, matched-filtering, and synchronization, the receiver observes

$$\mathbf{y}_{\mathrm{R2}} = \mathcal{H}_{\mathrm{XR}}\mathcal{F}\mathcal{H}_{\mathrm{TX}}\mathbf{x}_{\mathrm{T}} + \mathcal{H}_{\mathrm{XR}}\mathcal{F}\mathbf{n}_{\mathrm{X}} + \mathbf{n}_{\mathrm{R2}}. \tag{2.4}$$

This chapter considers the following two configurations: i) the two-hop channel (TH-no direct link) and ii) the relay channel (RC-with the direct link). Let $(\mathrm{T}) \in \{(\mathrm{TH}), (\mathrm{RC})\}$ represent the transmission strategy. The signal used for detection, denoted as $\mathbf{y}_{(\mathrm{T})}$, is obtained from (2.1) and (2.5). In particular, for the two-hop channel

$$\mathbf{y}_{(\mathrm{TH})} = \underbrace{\mathcal{H}_{\mathrm{XR}}\mathcal{F}\mathcal{H}_{\mathrm{TX}}}_{\mathcal{H}_{(\mathrm{TH})}}\mathbf{x}_{\mathrm{T}} + \underbrace{\mathcal{H}_{\mathrm{XR}}\mathcal{F}\mathbf{n}_{\mathrm{X}} + \mathbf{n}_{\mathrm{R2}}}_{\mathbf{n}_{(\mathrm{TH})}}. \tag{2.5}$$

While the stacked received signal for the relay channel is given by

$$\mathbf{y}_{(\mathrm{RC})} = \underbrace{\begin{pmatrix} \mathcal{H}_{\mathrm{TX}} \\ \mathcal{H}_{(\mathrm{TH})} \end{pmatrix}}_{\mathcal{H}_{(\mathrm{RC})}}\mathbf{x}_{\mathrm{T}} + \underbrace{\begin{pmatrix} \mathbf{n}_{\mathrm{R1}} \\ \mathbf{n}_{(\mathrm{TH})} \end{pmatrix}}_{\mathbf{n}_{(\mathrm{RC})}}. \tag{2.6}$$

A linear receive filter matrix $\mathcal{G}_{(\mathrm{T})}$ is applied to $\mathbf{y}_{(\mathrm{T})}$. To support linear detection, $\mathcal{H}_{(\mathrm{T})}$ must have at least as many rows as columns, i.e., $m_{\mathrm{T}} \leq N_{\mathrm{R}}$ (and $N_{\mathrm{T}} \leq N_{\mathrm{R}}$ for full spatial multiplexing). There are two popular types of linear receive filters: minimum mean squared error (MMSE) receive filters and ZF receive filters. MMSE receive filters slightly outperform ZF receive filters in terms of VSER performance. On the contrary, ZF receive filters have a complexity advantage and allow for more tractable analysis [83]. Note, however, that their diversity orders are the same. In this chapter, for tractability, this chapter focuses on only ZF receive filters. Because the equalized streams $\mathcal{G}_{(\mathrm{T})}\mathbf{y}_{(\mathrm{T})}$ are detected independently, the VSER is a function of post-processing SNR of each stream, denoted as $\mathrm{SNR}_{(\mathrm{T}),k}$ for $k = 1, 2, \cdots, m_{\mathrm{T}}$.

Having the relay filter design and all constituent channel information, the receiver determines the best mode. The selected mode index is sent back to the transmitter and relay via low-rate feedback channels. I assume the feedback channels are error-free (since the feedback channels usually have more coding) and zero-delay. In practice, however, signal processing and propagation result in feedback delays. This causes mismatching between the channels used for selecting a mode and the channels to which the mode is actually applied. Section 2.6 presents the simulation results on the impact of delays in feedback channels on the VSER performance of some of the proposed algorithms. The analysis on the impact of delays and the methods to compensate for delays are left for future work.

Each mode of operation is identified by the quadruple $(m_{\mathrm{T}}, p, m_{\mathrm{X}}, q)$, which I denote as $\omega \in \Omega$, where Ω is the set of all supported modes. I use the SNR-scaled singular values $\lambda_k(\mathbf{H}_{\mathrm{TR}}) = \sqrt{\gamma_{\mathrm{T}}}\sigma_k(\mathbf{H}_{\mathrm{TR}})$, $\lambda_k(\mathbf{H}_{\mathrm{TX}}) = \sqrt{\gamma_{\mathrm{T}}}\sigma_k(\mathbf{H}_{\mathrm{TX}})$ and $\lambda_k(\mathbf{H}_{\mathrm{XR}}) = \sqrt{\gamma_{\mathrm{X}}}\sigma_k(\mathbf{H}_{\mathrm{XR}})$, where $\gamma_{\mathrm{T}} = E_{\mathrm{T}}/\sigma^2$, $\gamma_{\mathrm{X}} = E_{\mathrm{X}}/\sigma^2$, $\sigma_k(\mathbf{H})$ is the k-th singular value of $\mathbf{H}$, and $\sigma_{\max}(\mathbf{H}) = \sigma_1(\mathbf{H}) \geq \cdots \geq \sigma_k(\mathbf{H}) \geq \cdots \geq \sigma_{\min}(\mathbf{H})$.

Remark 2.2.1. This remarks summarizes the key assumptions in this chapter and their justification.

- *Assumption 2.1: I assume a multiple-antenna AF relay aids data transmission from the transmitter to the receiver.* Multiple antennas are needed for simultaneously transmitting multiple data streams.

- *Assumption 2.2: The relay cannot transmit and receive at the same time.* This means I consider only half-duplex relays since they are more practical than full-duplex relays.

- *Assumption 2.3: The transmitter is silent in the second stage of the transmission procedure.* This requires non-orthogonal transmission protocols. Although the transmitter may send an extra message in the second stage to increase spectral efficiency in high SNR, the gain is not realized in low SNR.

- *Assumption 2.4: Perfect and instantaneous information of all channels is available at the receiver.* The receiver itself measures the channels from the transmitter and relay. The relay measures the channel from the transmitter to itself and then feeds forward to the receiver. My results show significant vector symbol error rate gains can be obtained under the ideal CSI assumption. They provide a bench mark for the papers that develop techniques with a more practical CSI assumption.

- *Assumption 2.5: The total number of bits transmitted from the transmitter to the receiver in a symbol vector is fixed.* This means that the reliability of transmission is of most interest in this paper.

- *Assumption 2.6: The data streams at the transmitter convey the same number of bits.* This corresponds to spatial multiplexing, thus avoiding the need of complicated bit allocation among data streams.

- *Assumption 2.7: The channels are frequency-flat, slowly-varying, and block-fading.* For example, the results can be used for a single carrier of MIMO OFDM AF relay systems.

- *Assumption 2.8: I assume that antenna selection is used at for transmission at both the transmitter and the receiver.* Antenna selection allows for the use of fewer RF chains than the number of antennas as negligible performance losses.

- *Assumption 2.9: The receiver uses the linear ZF receive filter.* This assumption is needed for making analysis tractable. Although this type of linear receive filter may perform worse than the linear MMSE receive filter in low SNR, there is little difference in their performance at high SNR.

- *Assumption 2.10: Received signals at the relay and receiver are corrupted by additive, circularly symmetric, complex spatially white Gaussian noise. Uncoordinated interference, if any, is treated as additive Gaussian noise.* This means that I assume uncoordinated interferers use independent random Gaussian codebooks, i.e., Gaussian signaling. Non-Gaussian interference is out of the scope of this dissertation.

- *Assumption 2.11: There exist error-free and zero-delay feedback channels from the receiver to the transmitter and relay for the receiver sends the indices of the selected antennas back to the transmitter and relay.* This assumption can be justified when the channels vary slow enough thus delay has negligible effects and feedback information is sent by using error-control coding and low-level modulation schemes.

2.3 VSER Analysis and Relay Filter Design

This section provides the VSER analysis of a general mode $\omega \in \Omega$ for the two-hop channel and relay channel. It also discusses the relay filter's design.

2.3.1 VSER Analysis for a General Mode of Operation

The VSER is a function of $\mathrm{SNR}_{(\mathrm{T}),k}(\omega)$ for $k = 1, 2, \cdots, m_{\mathrm{T}}$. Let $\mathbf{R}_{(\mathrm{T})}$ denote the covariance matrix of $\mathbf{n}_{(\mathrm{T})}$. For ZF receive filters [86]

$$\mathrm{SNR}_{(\mathrm{T}),k}(\omega) = \frac{E_{\mathrm{T}}/m_{\mathrm{T}}}{\left[\mathcal{H}^{*}_{(\mathrm{T})}\mathbf{R}_{(\mathrm{T})}\mathcal{H}_{(\mathrm{T})}\right]^{-1}_{kk}}. \tag{2.7}$$

The minimum post-processing SNR over m_{T} streams is denoted as $\mathrm{SNR}_{(\mathrm{T})} \triangleq \min_{1\le k\le m_{\mathrm{T}}} \mathrm{SNR}_{(\mathrm{T}),k}(\omega)$. Because closed-form expressions for the exact VSER are challenging to derive due to their complex mathematical structure, I propose to use the nearest neighbor upper bound (NNUB) [67, 86] to derive the closed-form bounds. Just like point-to-point MIMO links [83], the bounds can be used for the relative comparison of the VSER of supported modes to find the mode that is most likely to deliver the lowest VSER. In this chapter, the same method is applied to MIMO relay links. Let $d_{\min}(\omega)$ be the minimum distance and $N_e(\omega)$ be the average number of nearest neighbors of the transmit constellation [67]. The NNUB bound on the VSER is given by [83, 86]

$$P_{e,(\mathrm{T})}(\omega) = 1 - \left[1 - N_e(\omega) Q\left(\sqrt{\mathrm{SNR}_{(\mathrm{T})}(\omega)\frac{d^2_{\min}(\omega)}{2}}\right)\right]^{m_{\mathrm{T}}}. \tag{2.8}$$

The minimum stream post-processing SNR for full spatial multiplexing is defined as $\mathrm{SNR}^{(m)}_{(\mathrm{T})} \triangleq \min_{1\le k\le m_{\mathrm{T}}} \mathrm{SNR}_{(\mathrm{T}),k}(\omega)$. The corresponding NNUB bound on the VSER, which is denoted as $P^{(m)}_{e,(\mathrm{T})}$, is obtained by substituting $\omega \equiv (N_{\mathrm{T}}, 1, N_{\mathrm{X}}, 1)$ into (2.8). For single-stream transmissions, a mode is represented by $\omega \equiv (1, p, 1, q)$, where $1 \le p \le N_{\mathrm{T}}$ and $1 \le q \le N_{\mathrm{X}}$. Note that $\mathbf{W}_{\mathrm{T},(1,p)}$ and $\mathbf{W}_{\mathrm{X},(1,q)}$ are simply columns of the relevant identity matrices.

The post-processing SNR is

$$\mathrm{SNR}_{(\mathrm{T})}(1,p,1,q) = E_{\mathrm{T}}\left\|\mathbf{h}^{*}_{(\mathrm{T})}\left(\mathbf{R}^{(s)}_{(\mathrm{T})}\right)^{-1}\mathbf{h}_{(\mathrm{T})}\right\|, \tag{2.9}$$

where $\mathbf{h}_{(\mathrm{T})}$ is the effective end-to-end column channel. The maximum post-processing SNR over all single-stream transmission modes is defined as $\mathrm{SNR}^{(s)}_{(\mathrm{T})} \triangleq \max_{1\le p\le N_{\mathrm{T}},1\le q\le N_{\mathrm{X}}} \mathrm{SNR}_{(\mathrm{T})}(1,p,1,q)$. The NNUB bound on the VSER for full selection diversity, which is denoted as $P^{(s)}_{e,(\mathrm{T})}$, is obtained by substituting $\mathrm{SNR}^{(s)}_{(\mathrm{T})}$ into (2.8).

2.3.1.1 Post-Processing Stream SNR for the Two-Hop Channel

The two-hop channel is not a simple cascade of two conventional point-to-point MIMO links (i.e., the transmitter-relay and relay-receiver channels). Instead, the carefully designed relay filter creates an intelligent cascade of the two matrix channels. In addition to $\mathbf{n}_{\mathrm{R2}}$, the received signal is corrupted by $\mathbf{n}_{\mathrm{X}}$, which is amplified by $\mathcal{F}$ and then forwarded via $\mathcal{H}_{\mathrm{XR}}$. The combined noise $\mathbf{n}_{(\mathrm{TH})}$ has the covariance matrix $\mathbf{R}_{(\mathrm{TH})} = \sigma^2\{\mathcal{H}_{(\mathrm{TH})}[\mathcal{H}^{*}_{\mathrm{TX}}\mathcal{H}_{\mathrm{TX}}]^{-1}\mathcal{H}^{*}_{(\mathrm{TH})} + \mathbf{I}_{N_{\mathrm{R}}}\}$. The post-processing noise covariance matrix is $\mathbf{R}^{\mathrm{post}}_{(\mathrm{TH})} = \sigma^2\{[\mathcal{H}^{*}_{\mathrm{TX}}\mathcal{H}_{\mathrm{TX}}]^{-1} + [\mathcal{H}^{*}_{(\mathrm{TH})}\mathcal{H}_{(\mathrm{TH})}]^{-1}\}$, which is the sum of the local thermal noise at the receiver $\sigma^2[\mathcal{H}^{*}_{\mathrm{TX}}\mathcal{H}_{\mathrm{TX}}]^{-1}$ and the noise from the relay $\sigma^2[\mathcal{H}^{*}_{(\mathrm{TH})}\mathcal{H}_{(\mathrm{TH})}]^{-1}$. The post-processing SNR of the k-th stream is given by

$$\mathrm{SNR}_{(\mathrm{TH}),k}(\omega) = \frac{\gamma_{\mathrm{T}}/m_{\mathrm{T}}}{[\mathcal{H}^{*}_{\mathrm{TX}}\mathcal{H}_{\mathrm{TX}}]^{-1}_{kk} + [\mathcal{H}^{*}_{(\mathrm{TH})}\mathcal{H}_{(\mathrm{TH})}]^{-1}_{kk}}. \tag{2.10}$$

The minimum post-processing SNR over m_{T} streams in the two-hop channel is denoted as $\mathrm{SNR}_{(\mathrm{TH})}(\omega) \triangleq \min_{1\le k\le m_{\mathrm{T}}} \mathrm{SNR}_{(\mathrm{TH}),k}(\omega)$.

2.3.1.2 Post-Processing Stream SNR for the Relay Channel

To reach the receiver, the transmitter messages propagate via the following two paths: i) cascaded channels as in the two-hop channel and ii) parallel channels (i.e., the transmitter-receiver and transmitter-relay-receiver channels). Based on (2.6) and after some manipulation, I compute the post-processing SNR of the k-th stream as follows:

$$\mathrm{SNR}_{(\mathrm{RC}),k}(\omega) = \frac{\gamma_{\mathrm{T}}/m_{\mathrm{T}}}{\left\{\mathcal{H}_{\mathrm{TR}}^{*}\mathcal{H}_{\mathrm{TR}} + \sigma^{-2}\left(\mathbf{R}_{(\mathrm{TH})}^{\mathrm{post}}\right)^{-1}\right\}_{kk}^{-1}}. \tag{2.11}$$

In the relay channel, the minimum of the post-processing SNR values of m_{T} streams is denoted as $\mathrm{SNR}_{(\mathrm{RC})}(\omega) \triangleq \min_{1\le k\le m_{\mathrm{T}}} \mathrm{SNR}_{(\mathrm{RC}),k}(\omega)$.

2.3.2 Relay Filter Design

This subsection considers a suboptimal VSER minimization relay filter, which is given in [78] and aims to minimize the MSE of vector symbol estimation at the receiver. Although not directly minimizing the VSER, this relay filter design has the best VSER performance among the existing designs, including the mutual information maximization relay filters [141, 200] and uniform power allocation relay filter [154]. Although it is shown in [78] that the receiver should use the MMSE receive filter, I provide here a minor modification of the relay filter design in [78] when the ZF receive filter is used, which is in line with the system model considered in this chapter. Specifically, the MSE of vector symbol estimation is given as

$$J(\mathcal{F}) = \mathbb{E}[\|\hat{\mathbf{x}}_{\mathrm{T}} - \mathbf{x}_{\mathrm{T}}]\|^2] = \mathrm{tr}\{[\mathcal{H}_{\mathrm{TX}}^{*}\mathcal{F}\mathcal{H}_{\mathrm{TX}}]^{-1}\}, \tag{2.12}$$

where $\hat{\mathbf{x}}_{\mathrm{T}}$ is the estimated symbol vector. The relay filter is found by solving the MSE minimization problem, where the objective function is given in (2.12), subject to the power constraint at the relay in (2.3). I perform the singular value decomposition of $\mathcal{H}_{\mathrm{TX}}$ and $\mathcal{H}_{\mathrm{XR}}$, i.e., $\mathcal{H}_{\mathrm{TX}} = \mathcal{U}_{\mathrm{TX}}\Sigma_{\mathrm{TX}}\mathcal{V}_{\mathrm{TX}}^*$ and $\mathcal{H}_{\mathrm{XR}} = \mathcal{U}_{\mathrm{XR}}\Sigma_{\mathrm{XR}}\mathcal{V}_{\mathrm{XR}}^*$. Following the same proof as in [78], the relay filter must have the following form $\mathcal{F} = \mathcal{V}_{\mathrm{R}}\Lambda_{\mathrm{X}}\mathcal{U}_{\mathrm{TX}}^*$, where $\Lambda_{\mathrm{X}} = \mathrm{diag}(\lambda_1(\mathcal{F}), \lambda_2(\mathcal{F}), \cdots, \lambda_{m_{\mathrm{T}}}(\mathcal{F}))$ is the power allocation matrix. Let n_{TXR} be the maximum value of k, $1 \le k \le m_{\mathrm{T}}$ such that $\lambda_i(\mathcal{H}_{\mathrm{TX}}) > 0$ and $\lambda_i(\mathcal{H}_{\mathrm{XR}}) > 0$ for all $1 \le i \le k$. I define $a_i = \lambda_i^{-1}(\mathcal{H}_{\mathrm{TX}})\lambda_i^{-1}(\mathcal{H}_{\mathrm{XR}})$ and $b_i = \sqrt{\lambda_i^2(\mathcal{H}_{\mathrm{TX}})/m_{\mathrm{T}} + 1}$ for $i = 1, 2, \cdots, n_{\mathrm{TXR}}$. It follows that the optimal power allocation is

$$\lambda_i(\mathbf{F}) = \begin{cases} \sqrt{\frac{\gamma_{\mathrm{X}} a_i}{b_i \sum_{j=1}^{n_{\mathrm{TXR}}} a_j b_j}}, & \text{if } 1 \le i \le n_{\mathrm{TXR}}, \\ 0, & \text{if } n_{\mathrm{TXR}} < i \le m_{\mathrm{T}}. \end{cases} \tag{2.13}$$

2.4 Two-Hop Channel

When the direct link is weak, e.g., due to large path-loss and heavy shadowing, it can be neglected by the receiver. In this section, I propose a lower bound on the minimum post-processing SNR in a mode for the two-hop channel and then develop two-hop mode selection algorithms.

2.4.1 Performance Analysis of a Two-Hop Mode of Operation

A general mode of operation ω is considered in this subsection. A lower bound on $\mathrm{SNR}_{(\mathrm{TH})}(\omega)$ as a function of the eigenmodes of $\mathcal{H}_{\mathrm{TX}}$ and $\mathcal{H}_{\mathrm{XR}}$ is proposed in Theorem 2.4.1.

Theorem 2.4.1. *The minimum post-processing SNR among all streams is lower bounded by*

$$\mathrm{SNR}_{(\mathrm{TH})}(\omega) \geq \frac{1}{m_{\mathrm{T}}} \lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{(\mathrm{TH})}), \tag{2.14}$$

where $\lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{(\mathrm{TH})})$ *is defined as follows [with* $\lambda_{m_{\mathrm{T}}}(\mathcal{F})$ *given in (2.13)]*

$$\lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{(\mathrm{TH})}) \triangleq \frac{\lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{\mathrm{TX}})}{1 + \gamma_{\mathrm{X}} \lambda_{m_{\mathrm{T}}}^{-2}(\mathcal{F}) \lambda_{m_{\mathrm{T}}}^{-2}(\mathcal{H}_{\mathrm{XR}})}. \tag{2.15}$$

Proof. For any two matrices $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{B} \in \mathbb{C}^{n \times p}$ such that $\min\{m, n\} \leq \min\{n, p\}$, it follows from Ostrowskis theorem [89] that for any $1 \leq k \leq \min\{m, n\}$, there exists a positive real number θ_k such that $\sigma_{\min}(\mathbf{B}^*\mathbf{B}) \leq \theta_k \leq \sigma_{\max}(\mathbf{B}^*\mathbf{B})$ and $\sigma_k(\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{B}) = \theta_k \sigma_k(\mathbf{A}^*\mathbf{A})$. Note that $\sigma_k(\mathbf{A}^*\mathbf{A}) \geq 0$ because $\mathbf{A}^*\mathbf{A}$ is Hermitian. Thus, the following inequalities hold for any k, $1 \leq k \leq \min\{m, n\}$

$$\sigma_{\min}(\mathbf{B}^*\mathbf{B})\sigma_k(\mathbf{A}^*\mathbf{A}) \leq \sigma_k(\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{B}) \leq \sigma_{\max}(\mathbf{B}^*\mathbf{B})\sigma_k(\mathbf{A}^*\mathbf{A}). \tag{2.16}$$

The following lower bound on $\lambda_{\min}(\mathcal{H}_{(\mathrm{TH})})$ is obtained by applying (2.16) twice on $\mathcal{H}_{(\mathrm{TH})} = \mathcal{H}_{\mathrm{TX}}^{*} \mathcal{F}^{*} \mathcal{H}_{\mathrm{R}}^{*} \mathcal{H}_{\mathrm{R}} \mathcal{F} \mathcal{H}_{\mathrm{TX}}$ and changing to the use of the scaled singular values

$$\lambda_{\min}^{2}(\mathcal{H}_{(\mathrm{TH})}) \geq \gamma_{\mathrm{X}}^{2} \lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{\mathrm{TX}}) \lambda_{m_{\mathrm{T}}}^{2}(\mathcal{F}) \lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{\mathrm{XR}}). \tag{2.17}$$

Using $\max\{x + y\} \leq \max x + \max y$ and $\max_{1 \leq k \leq \min\{m,n\}} [\mathbf{H}^*\mathbf{H}]_{kk}^{-1} \leq \sigma_{\min}^{-2}(\mathbf{H})$ for any $\mathbf{H} \in \mathbb{C}^{m \times n}$, I obtain

$$\gamma_{\mathrm{T}}^{-1} \max_{1 \leq k \leq m_{\mathrm{T}}} [(\mathcal{H}_{\mathrm{TX}}^{*} \mathcal{H}_{\mathrm{TX}})_{kk}^{-1} + (\mathcal{H}_{(\mathrm{TH})}^{*} \mathcal{H}_{(\mathrm{TH})})_{kk}^{-1}] \leq \lambda_{\min}^{-2}(\mathcal{H}_{\mathrm{TX}}) + \lambda_{\min}^{-2}(\mathcal{H}_{(\mathrm{TH})}).$$

Applying the inequality into (2.10), I finish the proof of Theorem 1. $\square$

This bound provides insights into how the spatial characteristics of $\mathcal{H}_{\mathrm{TX}}$ and $\mathcal{H}_{\mathrm{XR}}$ affect SNR performance. Since $m_{\mathrm{T}} \leq N_{\mathrm{R}}$ and $\mathcal{H}_{(\mathrm{TH})} \in \mathbb{C}^{N_{\mathrm{R}} \times m_{\mathrm{T}}}$, $\mathcal{H}_{(\mathrm{TH})}$ has m_{T} singular values. Temporarily neglecting the relay and treating $\mathcal{H}_{(\mathrm{TH})}$ as a point-to-point MIMO link, I use the results in [83] to show that $\lambda_{\min}(\mathcal{H}_{(\mathrm{TH})})$ plays the same role as the minimum SNR-scaled singular value of $\mathcal{H}_{(\mathrm{TH})}$ (with the scale factor $\sqrt{\gamma_{\mathrm{T}}}$). Therefore, the SNR performance of any mode is determined by the minimum singular value of $\mathcal{H}_{(\mathrm{TH})}$, which is given in (2.15).

The effective minimum singular value provides insights into the two-hop channel operation. Here I use the convention that $(1/\infty) = 0$ and $(1/0) = \infty$. If $\mathcal{H}_{\mathrm{TX}}$ is not full-rank, i.e., $\lambda_{\min}(\mathcal{H}_{\mathrm{TX}}) = 0$, then the first hop channel is not suited for transmitting simultaneously m_{T} streams since at least one stream gets lost in the null space. Similarly, if $\mathrm{rank}(\mathcal{H}_{\mathrm{XR}}) < m_{\mathrm{T}}$, i.e., $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{\mathrm{XR}}) = 0$, then the second hop cannot support m_{T} streams. In both cases, substituting either $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{\mathrm{TX}})$ or $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{\mathrm{XR}})$ into (2.15), then $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{(\mathrm{TH})}) = 0$, which means that the mode is also inappropriate to transmit m_{T} independent streams. This observation is intuitively correct. If any of the two hops cannot support m_{T} streams, then the cascade of the two hops - despite how intelligent it is - cannot either. Moreover, $m_{\mathrm{T}} \leq m_{\mathrm{X}}$ must hold to support m_{T} streams, i.e., there is no need to consider the modes with $m_{\mathrm{T}} > m_{\mathrm{X}}$.

2.4.1.1 Full Spatial Multiplexing Analysis

The mode for full spatial multiplexing is represented by $(N_{\mathrm{T}}, 1, N_{\mathrm{X}}, 1)$. The effective channel is $\mathcal{H}^{(m)}_{(\mathrm{TH})} = \mathbf{H}_{\mathrm{XR}}\mathbf{F}\mathbf{H}_{\mathrm{TX}}$, where $\mathbf{F}$ is the corresponding relay filter. The minimum scaled singular value of $\mathcal{H}^{(m)}_{(\mathrm{TH})}$ is defined as

$$\lambda^2_{N_{\mathrm{T}}}\left(\mathcal{H}^{(m)}_{(\mathrm{TH})}\right) \triangleq \frac{\lambda^2_{N_{\mathrm{T}}}(\mathbf{H}_{\mathrm{TX}})}{1+\gamma_{\mathrm{X}}\lambda^{-2}_{N_{\mathrm{T}}}(\mathbf{F})\lambda^{-2}_{N_{\mathrm{T}}}(\mathbf{H}_{\mathrm{XR}})}. \tag{2.18}$$

It follows from Theorem 2.4.1 that $\mathrm{SNR}^{(m)}_{(\mathrm{TH})} \geq (1/N_{\mathrm{T}})\lambda^2_{N_{\mathrm{T}}}(\mathcal{H}^{(m)}_{(\mathrm{TH})})$. The SNR performance of two-hop full spatial multiplexing systems is determined by the minimum SNR-scaled singular value of the effective two-hop channel.

Now I assume both hops can support N_{T} stream transmission, i.e., $n_{\mathrm{TXR}} = N_{\mathrm{T}} \leq \min\{N_{\mathrm{X}}, N_{\mathrm{R}}\}$ and $\lambda_i(\mathbf{H}_{\mathrm{TX}}) > 0$ and $\lambda_i(\mathbf{H}_{\mathrm{XR}}) > 0$ for all $1 \leq i \leq N_{\mathrm{T}}$. It follows from (2.13) that b_i/a_i increases in both $\lambda_i(\mathbf{H}_{\mathrm{TX}})$ and $\lambda_i(\mathbf{H}_{\mathrm{XR}})$ for all $1 \leq i \leq n_{\mathrm{T}}$. Because the singular values are in the descending order, $b_{N_{\mathrm{T}}}/a_{N_{\mathrm{T}}} \leq b_i/a_i$ for all $1 \leq i \leq N_{\mathrm{T}}$, and hence $\gamma_{\mathrm{X}}\lambda^{-2}_{N_{\mathrm{T}}}(\mathbf{F}) \leq (1/N_{\mathrm{T}})\sum_{i=1}^{N_{\mathrm{T}}}\lambda_i^2(\mathbf{H}_{\mathrm{TX}}) + N_{\mathrm{T}}$. The following two lower bounds on $\lambda^2_{N_{\mathrm{T}}}(\mathcal{H}^{(m)}_{(\mathrm{TH})})$ are obtained by substituting this inequality into (2.18) and using the equalities $\sum_{i=1}^{N_{\mathrm{T}}}\lambda_i^2(\mathbf{H}_{\mathrm{TX}}) \leq N_{\mathrm{T}}\lambda^2_{\max}(\mathbf{H}_{\mathrm{TX}})$ and $\lambda_{N_{\mathrm{T}}}(\mathbf{H}_{\mathrm{XR}}) \geq \lambda_{\min}(\mathbf{H}_{\mathrm{XR}})$

$$\begin{aligned}
\lambda^2_{N_{\mathrm{T}}}(\mathcal{H}^{(m)}_{(\mathrm{TH})}) &\geq \frac{\lambda^2_{\min}(\mathbf{H}_{\mathrm{TX}})\lambda^2_{N_{\mathrm{T}}}(\mathbf{H}_{\mathrm{XR}})}{\frac{1}{N_{\mathrm{T}}}\sum_{i=1}^{N_{\mathrm{T}}}\lambda_i^2(\mathbf{H}_{\mathrm{TX}}) + \lambda^2_{N_{\mathrm{T}}}(\mathbf{H}_{\mathrm{XR}}) + N_{\mathrm{T}}} && (2.19)\\
&\geq \frac{\lambda^2_{\min}(\mathbf{H}_{\mathrm{TX}})\lambda^2_{\min}(\mathbf{H}_{\mathrm{XR}})}{\lambda^2_{\max}(\mathbf{H}_{\mathrm{TX}}) + \lambda^2_{\min}(\mathbf{H}_{\mathrm{XR}}) + N_{\mathrm{T}}}. && (2.20)
\end{aligned}$$

Note that (2.19) is the minimum SNR-scaled singular value of the effective two-hop channel when using the uniform power allocation at the relay while the proposed algorithm uses the optimal one.

2.4.1.2 Single-Stream Antenna Selection Diversity Analysis

The effective channel is $\mathbf{h}_{(\mathrm{TH}),(1,p,1,q)} = \alpha_{\mathrm{X}}^{(s)}\|\mathbf{h}_{\mathrm{TX},p}\|\mathbf{h}_{\mathrm{XR},q}$, where $\alpha_{\mathrm{X}}^{(s)} = \gamma_{\mathrm{X}}^{1/2}(\gamma_{\mathrm{T}}\|\mathbf{h}_{\mathrm{TX},p}\|^2+1)^{-1/2}$ is the amplification factor that satisfies the power constraint at the relay. From (2.9), the post-processing SNR of the single-stream transmissions is given by

$$\begin{aligned}\mathrm{SNR}_{(\mathrm{TH}),1}(1,p,1,q) &= \left(\frac{1}{\gamma_{\mathrm{T}}}\|\mathbf{h}_{\mathrm{TX},p}\|^{-2}+\frac{1}{\gamma_{\mathrm{X}}}\|\mathbf{h}_{\mathrm{XR},p}\|^{-2}\right.\\ &\left.+\frac{1}{\gamma_{\mathrm{T}}\gamma_{\mathrm{X}}}\|\mathbf{h}_{\mathrm{TX},p}\|^{-2}\|\mathbf{h}_{\mathrm{XR},p}\|^{-2}\right)^{-1}, \qquad (2.21)\end{aligned}$$

which depends on p and q. The most common antenna selection algorithm for this case is to select p and q that maximize the post-processing SNR [12]. From (2.21), $\mathrm{SNR}_{(\mathrm{TH}),1}(1,p,1,q)$ is strictly increasing in both $\|\mathbf{h}_{\mathrm{TX},p}\|^{-2}$ (which depends only on p) and $\|\mathbf{h}_{\mathrm{XR},p}\|^{-2}$ (which depends only on q). Thus the minimization of $\mathrm{SNR}_{(\mathrm{TH}),1}(1,p,1,q)$ can be achieved by maximizing the channel gains $\|\mathbf{h}_{\mathrm{TX},p}\|^{-2}$ and $\|\mathbf{h}_{\mathrm{XR},p}\|^{-2}$ independently. This observation helps reduce computational complexity significantly. I denote $p^* \triangleq \arg\max_{1\le p\le N_{\mathrm{T}}}\|\mathbf{h}_{\mathrm{TX},p}\|$, $q^* \triangleq \arg\max_{1\le p\le N_{\mathrm{T}}}\|\mathbf{h}_{\mathrm{XR},q}\|$, and $\mathrm{SNR}_{(\mathrm{TH})}^{(s)} \triangleq \mathrm{SNR}_{(\mathrm{TH}),1}(1,p^*,1,q^*)$. An upper bound on $\mathrm{SNR}_{(\mathrm{TH})}^{(s)}$ is

$$\begin{aligned}\mathrm{SNR}_{(\mathrm{TH})}^{(s)} &\le \big[\lambda_{\max}(\mathbf{H}_{\mathrm{TX}})^{-2}+\lambda_{\max}(\mathbf{H}_{\mathrm{XR}})^{-2}\\ &+\lambda_{\max}(\mathbf{H}_{\mathrm{TX}})^{-2}\lambda_{\max}(\mathbf{H}_{\mathrm{XR}})^{-2}\big]^{-1}, \qquad (2.22)\end{aligned}$$

where (2.22) is obtained by $\min_{1\le k\le \min\{m,n\}}\|\mathbf{h}_k\| \le \sigma_{\max}(\mathbf{H})$ for any $\mathbf{H}\in\mathbb{C}^{m\times n}$ [89]. The upper bound on the SNR performance of two-hop selection diversity systems in (2.22) is an increasing function of both $\lambda_{\max}(\mathbf{H}_{\mathrm{TX}})^{-2}$ and

$\lambda_{\max}(\mathbf{H}_{\mathrm{XR}})^{-2}$. The result is expected since the upper bound on the SNR performance of the point-to-point MIMO selection diversity in [83] increases strictly with the square of its maximum singular value. This bound thus acts like the maximum SNR-scaled singular value of $\mathcal{H}_{(\mathrm{TH})}^{(s)}$. If I define $\lambda_{\max}^2(\mathcal{H}_{(\mathrm{TH})}^{(s)})$ as the bound in (2.22), or equivalently

$$\lambda_{\max}^2\left(\mathcal{H}_{(\mathrm{TH})}^{(s)}\right) \triangleq \frac{\lambda_{\max}^2(\mathbf{H}_{\mathrm{TX}})\lambda_{\max}^2(\mathbf{H}_{\mathrm{XR}})}{\lambda_{\max}^2(\mathbf{H}_{\mathrm{TX}}) + \lambda_{\max}^2(\mathbf{H}_{\mathrm{XR}}) + 1}, \tag{2.23}$$

then (2.22) can be rewritten as $\mathrm{SNR}_{(\mathrm{TH})}^{(s)} \leq \lambda_{\max}^2(\mathcal{H}_{(\mathrm{TH})}^{(s)})$.

2.4.2 Dualmode Two-Hop Transmission: Full Spatial Multiplexing versus Full Selection Diversity

This section presents a sufficient condition that full spatial multiplexing causes a lower VSER than full selection diversity. This forms the principle of dynamic dualmode algorithms for the two-hop channel.

2.4.2.1 VSER-based Selection

The first dualmode selection criterion is based on $P_{e,(\mathrm{TH})}^{(m)}$ and $P_{e,(\mathrm{TH})}^{(s)}$. The NNUB bounds are determined based on $\mathrm{SNR}_{(\mathrm{TH})}^{(m)}$ and $\mathrm{SNR}_{(\mathrm{TH})}^{(s)}$, respectively, using (2.8) as discussed in Subsection 2.3.1. The selection criterion is simply to choose the mode with the lower NNUB bound.

Two-Hop Selection Criterion 1 - VSER-based Selection: Choose spatial multiplexing if

$$P_{e,(\mathrm{TH})}^{(m)} < P_{e,(\mathrm{TH})}^{(s)};$$

otherwise, choose selection diversity transmission with the best antennas at the transmitter and relay.

Given the SNR values, it follows from (2.8) that the main computational load of the algorithm is involved with the implementation of the Q-function. Although there exist solutions for this issue, e.g., to use a look-up table or a polynomial approximation method [81], the computational complexity is still high. More importantly, the algorithm does not reveal much about the role of the quality of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$ in the mode selection. This motivates the development of lower-complexity algorithms that might provide more insights into how to select the mode based on knowledge of the constituent channels.

2.4.2.2 Post-Processing SNR-based Selection

The following observations help to avoid the use of the Q-function. First, $1-(1-x)^n \approx nx$ for small x and an arbitrary positive integer n. This is obtained by applying the binomial theorem and by eliminating the terms with high order of x. This approximation transforms the expressions of $P_{e,(\mathrm{TH})}^{(m)}$ and $P_{e,(\mathrm{TH})}^{(s)}$ into the products of the number of the nearest neighbors and the Q-function. Next, since the overall rate is fixed, the numbers of points in the vector constellations are the same, thus the term of the number of nearest neighbors in the relative comparison of VSERs can be neglected. Finally, since the Q-function decreases with increasing arguments, I propose the following selection criterion.

Two-Hop Selection Criterion 2 - Post-Processing SNR-based Selection:

Choose spatial multiplexing if

$$d^2_{\min}(N_{\mathrm{T}}, R)\mathrm{SNR}^{(m)}_{(\mathrm{TH})} < d^2_{\min}(1, R)\mathrm{SNR}^{(s)}_{(\mathrm{TH})};$$

otherwise, choose selection diversity transmission with the best antennas at the transmitter and relay.

This algorithm shows how the effective end-to-end channels (through the post-processing SNR values) and number of data streams (through the minimum constellation distances) affect dualmode selection. Specifically, while the VSER performance of full spatial multiplexing depends significantly on the post-processing SNR of the worst stream, the VSER performance of full selection diversity is a function of the columns with the largest norm of the original channels on two hops. Furthermore, the two modes have different minimum constellation distances. Full spatial multiplexing, which sends only R/N_{T} bits per stream, uses a constellation with a larger minimum constellation distance than full selection diversity, which sends all R bits in one stream.

2.4.2.3 Condition Number-based Selection

I now develop a selection criterion that is based directly on the spatial characteristics of the original constituent channel realizations $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$. This follows the idea of a dualmode algorithm that is based on the regular condition number of conventional MIMO links [83]. In particular, I propose the concept of the cascade condition number of the two-hop channel as the ratio of the maximum singular value [given in (2.15)] over the minimum singular value [given in (2.23)]

$$\kappa_{(\mathrm{TH})} \triangleq \frac{\lambda_{\max}\big(\mathcal{H}^{(m)}_{(\mathrm{TH})}\big)}{\lambda_{\min}\big(\mathcal{H}^{(s)}_{(\mathrm{TH})}\big)}. \tag{2.24}$$

The cascade condition number is in fact a function of the spatial characteristics of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$ via their singular values. The lower bound of $\lambda_{\min}\big(\mathcal{H}^{(s)}_{(\mathrm{TH})}\big)$ given in (2.20) is used to rewrite $\kappa_{(\mathrm{TH})}$ as

$$\kappa^2_{(\mathrm{TH})} \approx \kappa^2_{\mathrm{TX}}\kappa^2_{\mathrm{XR}}\frac{\lambda^2_{\max}(\mathbf{H}_{\mathrm{TX}}) + \lambda^2_{\min}(\mathbf{H}_{\mathrm{XR}}) + N_{\mathrm{T}}}{\lambda^2_{\max}(\mathbf{H}_{\mathrm{TX}}) + \lambda^2_{\min}(\mathbf{H}_{\mathrm{XR}}) + 1}, \tag{2.25}$$

where $\kappa_{\mathrm{TX}} = \lambda_{\max}(\mathbf{H}_{\mathrm{TX}})/\lambda_{\min}(\mathbf{H}_{\mathrm{TX}})$ and $\kappa_{\mathrm{XR}} = \lambda_{\max}(\mathbf{H}_{\mathrm{XR}})/\lambda_{\min}(\mathbf{H}_{\mathrm{XR}})$ are the regular condition numbers of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$, respectively. The approximation in (2.25) becomes precise when $N_{\mathrm{T}} = \min\{N_{\mathrm{X}}, N_{\mathrm{R}}\}$ and $\kappa_{\mathrm{TX}} = 1$ (such that $\lambda_{N_{\mathrm{T}}}(\mathbf{H}_{\mathrm{XR}}) = \lambda_{\min}(\mathbf{H}_{\mathrm{XR}})$ and $\sum_{i=1}^{N_{\mathrm{T}}} \lambda_i^2(\mathbf{H}_{\mathrm{TX}}) = N_{\mathrm{T}}\lambda^2_{\max}(\mathbf{H}_{\mathrm{TX}})$. It can be shown that $\kappa_{(\mathrm{TH})}$ has the same characteristics as the regular condition number of the point-to-point MIMO channel, such as $\kappa_{(\mathrm{TH})} > 1$ and has no upper bound. Indeed, when any of the constituent channels is singular (full spatial multiplexing cannot be supported by that point-to-point channel), the cascade condition number is infinite, thus the two-hop channel is not suited for full spatial multiplexing. Moreover, $\kappa_{(\mathrm{TH})} > \kappa_{\mathrm{TX}}$, this means that even the optimal relay transceiver design cannot help improve the channels suitability to full spatial multiplexing.

Note that (2.18) provides a lower bound on the SNR performance of two-hop full spatial multiplexing and (2.21) gives an upper bound on the SNR performance of two-hop full selection diversity. Since these bounds are in opposite directions, they provide a sufficient condition for full spatial multiplexing to outperform full selection diversity. In other words, spatial multiplexing is chosen only if its worst minimum SNR is better than the best maximum SNR in full selection diversity.

Two-Hop Selection Criterion 3 - Condition Number-based Selection: Choose spatial multiplexing if

$$\kappa^2_{(\mathrm{TH})} < \frac{1}{\sqrt{N_\mathrm{T}}} \frac{d_{\min}(N_\mathrm{T}, R)}{d_{\min}(1, R)};$$

otherwise, choose selection diversity transmission with the best antennas at the transmitter and relay.

First, this selection is biased to full selection diversity and is a counterpart of *Selection Criterion 3* for point-to-point MIMO systems in [83]. Furthermore, this algorithm simplifies the computational load since just singular value decompositions for $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$ are needed. Lastly, the selection criterion is able to relate the spatial characteristics of the constituent channels, represented by the cascade condition number, directly to the minimum constellation distances and data rates.

2.4.3 Multimode Two-Hop Transmission

To support dualmode transmission, the numbers of transmit RF chains at the transmitter and relay must be equal to the numbers of transmit antennas, i.e., $M_\mathrm{T} = N_\mathrm{T}$ and $M_\mathrm{X} = N_\mathrm{X}$. In this section, to avoid these constraints and to allow more freedom for antenna selection, I develop multimode two-hop algorithms. In particular, multimode transmission allows the approaches that are in-between the two extreme modes supported by dualmode transmission. In other words, they support different combinations of m_{T} antennas at the transmitter, $m_{\mathrm{T}} \in \mathcal{M}_{\mathrm{T}}$, and m_{X} antennas at the relay, $1 \leq m_{\mathrm{X}} \leq M_{\mathrm{X}}$. It is worth to emphasize that multimode transmission works well for any $M_{\mathrm{T}} \leq N_{\mathrm{T}}$ and $M_{\mathrm{X}} \leq N_{\mathrm{X}}$. As discussed in Subsection 2.4.1, to support m_{T} streams, there is no need for considering $m_{\mathrm{T}} > m_{\mathrm{X}}$ in multimode transmission. Finally, the discussion on dualmode transmission still applies for multimode transmission, but with a larger pool of mode candidates.

2.4.3.1 VSER-based Multimode Antenna Selection

Since it is almost impossible to determine the exact conditional VSERs, I propose the first selection criterion based on their closed-form NNUB bounds.

Two-Hop Selection Criterion 4 - VSER-based Selection: Choose ω^* that solves

$$\arg\min_{\omega\in\Omega} 1 - \left[1 - N_e Q\left(\sqrt{\mathrm{SNR}_{(\mathrm{TH})}(\omega)\frac{d_{\min}^2(\omega)}{2}}\right)\right]^{m_{\mathrm{T}}}. \tag{2.26}$$

As an extension for *Two-hop Selection Criterion 1*, the algorithm provides few insights into how mode selection depends on the quality of the constituent channels. Furthermore, the implementation of this algorithm requires the determination of the post-processing SNR for each stream in each antenna selection policy, the computation of the Q-function for each policy, and an exhaustive search over all possible modes. These are the main sources of high computational complexity and demand a large storage memory for temporary values, making the algorithm less attractive in practice.

2.4.3.2 SNR-based Multimode Antenna Selection

Similar to the derivation of Two-hop Selection Criterion 2 in Section 2.4.2, I use an approximation of the NNUB bound, the assumption of fixed overall data rate, and a property of the Q-function (which decreases with increasing arguments). This results in the following selection criterion based on the post-processing stream SNR and minimum constellation distance.

Two-Hop Selection Criterion 5 - SNR-based Selection: Choose ω^* that solves

$$\arg\min_{\omega \in \Omega} d_{\min}^2(\omega)\mathrm{SNR}_{(\mathrm{TH})}(\omega). \tag{2.27}$$

Intuitively, with appropriate matched-filtering and precoding, the MIMO two-hop channel can be decomposed into m_{T} parallel subchannels whose gains are proportional to the singular values of the channel. Increasing m_{T} reduces the power allocated for each stream and leads to the use of subchannels with lower gains (which correspond to smaller singular values of channels on two hops). Nevertheless, for a fixed data rate R, as m_{T} increases, the number of bits carried on each stream $2^{R/m_{\mathrm{T}}}$ decreases, i.e., the constellation used for modulating the streams has fewer symbols and hence has a larger minimum constellation distance. In other words, the algorithm is based on the tradeoff between the minimum post-processing SNR over all streams and minimum constellation distance while changing the number of streams.

By avoiding the implementation of the Q-function, this algorithm has lower complexity than the VSER-based algorithm; however, it still requires the calculation of the post-processing SNR of all streams for all possible modes. Based on (2.10), I notice that the computation of the post-processing SNR for each mode is involved with the determination of the relay filter corresponding to each mode and several matrix operations such as matrix multiplication and matrix inverse. It has been shown in Theorem 2.4.1 that the SNR performance of any two-hop mode is determined by $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{(\mathrm{TH})})$, leading to a straightforward simplification of the *Two-hop Selection Criterion 5*, where $\mathrm{SNR}_{(\mathrm{TH})}(\omega)$ is replaced by a function of $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{(\mathrm{TH})})$. The computation of $\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{(\mathrm{TH})})$ needs only the determination of the relay filter, which includes already the computation of the m_{T}-th singular values of the effective channels on two hops. This simplified algorithm, however, still requires computing all the singular values of $\mathcal{H}_{(\mathrm{TH})}$ for all modes.

2.4.3.3 Eigenmode-based Multimode Antenna Selection

To simplify further multimode selection and to gain insights into how the spatial characteristics of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$ affect multimode transmission, I relate $\lambda^2_{m_{\mathrm{T}}}(\mathcal{H}_{(\mathrm{TH})})$ (using its lower bound in (2.19)) to the eigenmodes of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$. From Theorem 3 in [83], I have $\lambda^2_{m_{\mathrm{T}}}(\mathbf{H}_{\mathrm{TX}}) \geq \lambda^2_{m_{\mathrm{T}}}(\mathcal{H}_{\mathrm{TX}})$ and $\lambda^2_{m_{\mathrm{T}}}(\mathbf{H}_{\mathrm{XR}}) \geq \lambda^2_{m_{\mathrm{T}}}(\mathcal{H}_{\mathrm{XR}})$ for all $1 \leq m_{\mathrm{T}} \leq M_{\mathrm{T}}$, where the equalities are achieved when $\mathbf{W}_{\mathrm{T},(m_{\mathrm{T}},p)}$ consists of the m_{T} dominant right singular vectors of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{W}_{\mathrm{X},(m_{\mathrm{X}},q)}$ is a permutation of the m_{X} dominant right singular vectors of $\mathbf{H}_{\mathrm{XR}}$. By replacing $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{TX}})$ and $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{XR}})$ by $\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{TX}})$ and $\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}})$, respectively, in the right-hand side (RHS) of (2.19), I obtain an approximation for the lower bound on the minimum post-processing SNR over all streams as

$$\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{TH})}) \approx \frac{\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{TX}})\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}})}{\frac{1}{m_\mathrm{T}}\sum_{i=1}^{m_\mathrm{T}}\lambda_i^2(\mathbf{H}_{\mathrm{TX}}) + \lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}}) + m_\mathrm{T}}. \tag{2.28}$$

The beauty of the approximation in (2.28) is that it depends only on the number of streams m_T. This makes it possible to separate the choice of m_T from the choice of (p, m_X, q). How does the approximation depend on the other parameters (i.e., p, m_X and q)? The dependence is represented indirectly via the replacements of $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{TX}})$ and $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{XR}})$ by their corresponding upper bounds $\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{TX}})$ and $\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}})$. This means that given m_T, p should be found so that $\lambda_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{TX}}\mathbf{W}_{\mathrm{T},(m_\mathrm{T},p)})$ as close to $\lambda_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{TX}})$ as possible to improve the VSER performance. Similarly, (m_X, q) should be chosen jointly to maximize $\lambda_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}}\mathbf{W}_{\mathrm{X},(m_\mathrm{X},q)})$. Based on the observations, I propose the following selection criterion.

Two-Hop Selection Criterion 6 - Eigenmode-based Selection Criterion: The selection procedure is:

Step 1: Solve for

$$m_\mathrm{T}^* = \arg\max_{m_\mathrm{T}\in\mathcal{M}_\mathrm{T}} \frac{d^2_{\min}(m_\mathrm{T}, R)}{m_\mathrm{T}} \frac{\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{TX}})\lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}})}{\frac{1}{m_\mathrm{T}}\sum_{i=1}^{m_\mathrm{T}}\lambda_i^2(\mathbf{H}_{\mathrm{TX}}) + \lambda^2_{m_\mathrm{T}}(\mathbf{H}_{\mathrm{XR}}) + m_\mathrm{T}}$$

Step 2: Find (m_X^*, q^*) that solve

$$\arg\max_{m_\mathrm{T}^* \le m_\mathrm{X} \le M_\mathrm{X}, 1 \le q \le \binom{N_\mathrm{X}}{m_\mathrm{X}}} \lambda_{m_\mathrm{T}^*}(\mathbf{H}_{\mathrm{XR}}\mathbf{W}_{\mathrm{X},(m_\mathrm{X},q)})$$

Step 3: Choose p^* that solves

$$\arg\max_{1\leq p\leq\binom{N_{\mathrm{T}}}{m_{\mathrm{T}}^*}} \lambda_{m_{\mathrm{T}}^*}(\mathbf{H}_{\mathrm{TX}}\mathbf{W}_{\mathrm{T},(m_{\mathrm{T}}^*,p)}).$$

Regarding the implementation of the algorithm, step 1 requires only the singular value decompositions of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$ to determine the number of streams m_{T}^*. Next, step 2 involves the computation of the m_{T}^*-th singular value of the effective second-hop channel for $\sum_{m_{\mathrm{X}}=m_{\mathrm{T}}^*}^{M_{\mathrm{X}}}\binom{N_{\mathrm{X}}}{m_{\mathrm{X}}}$ modes [68]. In step 3, it is necessary to calculate the m_{T}^*-th singular value of the effective first-hop channel for at most $\binom{N_{\mathrm{T}}}{\lfloor N_{\mathrm{T}}/2\rfloor}$ modes. Note that even in the simplified version of the SNR-based algorithm, which is based on $\lambda_{m_{\mathrm{T}}}^2(\mathcal{H}_{(\mathrm{TH})})$, it is still necessary to compute the relay filter and hence the singular values of the corresponding effective channels on two hops for all $\left(\sum_{m_{\mathrm{T}}\in M_{\mathrm{T}}}\binom{N_{\mathrm{T}}}{m_{\mathrm{T}}}\right)(2^{N_{\mathrm{X}}}-1)$ possible modes. Therefore, the eigenmode-based multimode algorithm has much lower complexity than the other multimode algorithms.

An intuition about the eigenmode-based selection in case of low-rank constituent channels can be obtained. Let $r_{(\mathrm{TH})}=\min\{\mathrm{rank}(\mathbf{H}_{\mathrm{TX}}),\mathrm{rank}(\mathbf{H}_{\mathrm{XR}})\}$ be the minimum of the ranks of $\mathbf{H}_{\mathrm{TX}}$ and $\mathbf{H}_{\mathrm{XR}}$. There is no need to consider $m_{\mathrm{T}}>r_{(\mathrm{TH})}$. Because when $m_{\mathrm{T}}>r_{(\mathrm{TH})}$, the resulting end-to-end channel is nonsingular and hence cannot support transmissions of m_{T} independent streams.

2.5 Relay Channel

When the transmitter-receiver channel is good enough, the receiver can utilize the observations it receives in two stages for symbol detection. This section provides VSER performance analysis of the relay channel and then develops several antenna selection criteria.

2.5.1 Relay Channel Performance Analysis

Theorem 2.5.1 provides a lower bound on $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})})$, which is related to the eigenmodes of the effective constituent channels.

Theorem 2.5.1. *The minimum post-processing SNR over all streams of the relay channel satisfies*

$$\mathrm{SNR}_{(\mathrm{RC})}(\omega) \geq \lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})}), \tag{2.29}$$

where $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})})$ *is the minimum SNR-scaled singular value of* $\mathcal{H}_{(\mathrm{RC})}$ *and is defined as*

$$\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})}) \triangleq \lambda^2_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{TR}}) + \frac{\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{TX}})}{1 + \gamma_\mathrm{X}\lambda^{-2}_{m_\mathrm{T}}(\mathcal{F})\lambda^{-2}_{m_\mathrm{T}}(\mathcal{H}_{\mathrm{XR}})}. \tag{2.30}$$

Proof. Since $\mathbf{R}_{(\mathrm{TH})}$ is Hermitian, both $\mathbf{R}^{-1}_{(\mathrm{TH})}$ and $\mathcal{H}_{(\mathrm{TH})}\mathbf{R}^{-1}_{(\mathrm{TH})}\mathcal{H}^*_{(\mathrm{TH})}$ are Hermitian [89]. Note that

$$
\begin{aligned}
& \max_{1\le k\le m_{\mathrm{T}}} \left\{ \sigma^{-2}\mathcal{H}_{\mathrm{TR}}^{*}\mathcal{H}_{\mathrm{TR}} + \left[\mathcal{H}_{(\mathrm{TH})}^{\dagger}\mathbf{R}_{(\mathrm{TH})}(\mathcal{H}_{(\mathrm{TH})}^{*})^{\dagger} \right]^{-1} \right\}_{kk}^{-1} \\
\le\ & \sigma_{\max}\left(\sigma^{-2}\mathcal{H}_{\mathrm{TR}}^{*}\mathcal{H}_{\mathrm{TR}} + \left[\mathcal{H}_{(\mathrm{TH})}^{\dagger}\mathbf{R}_{(\mathrm{TH})}(\mathcal{H}_{(\mathrm{TH})}^{*})^{\dagger} \right]^{-1} \right) && (2.31) \\
\le\ & \left\{ \sigma^{-2}\sigma_{\min}(\mathcal{H}_{\mathrm{TR}}^{*}\mathcal{H}_{\mathrm{TR}}) + \sigma_{\min}\left(\left[\mathcal{H}_{(\mathrm{TH})}^{\dagger}\mathbf{R}_{(\mathrm{TH})}(\mathcal{H}_{(\mathrm{TH})}^{*})^{\dagger} \right]^{-1} \right) \right\}^{-1} && (2.32) \\
\le\ & \sigma^{2}\Big\{ \sigma_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{\mathrm{TR}}) \\
& \quad + \left[\sigma_{\max}([\mathcal{H}_{\mathrm{TX}}^{*}\mathcal{H}_{\mathrm{TX}}]^{-1}) + \sigma_{\max}\left([\mathcal{H}_{(\mathrm{TH})}^{*}\mathcal{H}_{(\mathrm{TH})}]^{-1} \right) \right]^{-1} \Big\}^{-1} && (2.33) \\
\le\ & E_{\mathrm{T}}\left[\lambda_{m_{\mathrm{T}}}(\mathcal{H}_{\mathrm{TR}}) + \frac{\lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{\mathrm{TX}})}{1+\gamma_{\mathrm{X}}\lambda_{m_{\mathrm{T}}}^{-2}(\mathcal{F})\lambda_{m_{\mathrm{T}}}^{-2}(\mathcal{H}_{\mathrm{XR}})} \right]^{-1}, && (2.34)
\end{aligned}
$$

where (2.31) results from $\max_{1\le k\le \min\{m,n\}}[\mathbf{H}^{*}\mathbf{H}]_{kk}^{-1} \le \sigma_{\min}^{-2}(\mathbf{H})$ for any $\mathbf{H} \in \mathbb{C}^{m\times n}$ [89]; (2.32) and (2.33) are obtained by applying the Weyl's inequality [89]; and (2.34) follows from (2.17). The result in Theorem 2.5.1 is obtained by substituting (2.33) into (2.11). □

Note that $\lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{(\mathrm{RC})}) = \lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{\mathrm{TR}}) + \lambda_{m_{\mathrm{T}}}^{2}(\mathcal{H}_{(\mathrm{TH})})$ for any given ω. This means that with appropriate signal processing at the receiver, the minimum post-processing SNR of the relay channel may be equal to the sum of the minimum post-processing SNR values of the direct channel and the two-hop channel. Thus, an efficient use of the direct channel helps improve the SNR performance; this is intuitively correct since it has more observations.

2.5.1.1 Full Spatial Multiplexing Analysis

Denote $\mathcal{H}_{\text{(RC)}}^{(m)} = \mathcal{H}_{\text{(RC)},(N_\text{T},1,N_\text{X},1)}$ as the effective full spatial multiplexing relay channel. From Theorem 2.5.1 for $\omega \equiv (N_\text{T}, 1, N_\text{X}, 1)$, a lower bound on $\text{SNR}_{\text{(RC)}}^{(m)}$ is

$$\text{SNR}_{\text{(RC)}}^{(m)} \geq \frac{1}{N_\text{T}} \lambda_{N_\text{T}}^2(\mathcal{H}_{\text{(RC)}}^{(m)}), \tag{2.35}$$

where $\lambda_{N_\text{T}}^2(\mathcal{H}_{\text{(RC)}}^{(m)})$ acts as the minimum scaled singular value of $\mathcal{H}_{\text{(RC)}}^{(m)}$. Similar to (2.30) and based on (2.20), it follows that

$$\begin{aligned}
\lambda_{N_\text{T}}^2(\mathcal{H}_{\text{(RC)}}^{(m)}) &\triangleq \lambda_{N_\text{T}}^2(\mathcal{H}_\text{TR}) + \frac{\lambda_{N_\text{T}}^2(\mathcal{H}_\text{TX})}{1 + \gamma_\text{X} \lambda_{N_\text{T}}^{-2}(\mathcal{F}) \lambda_{N_\text{T}}^{-2}(\mathcal{H}_\text{XR})} && (2.36)\\
&\geq \lambda_{\min}^2(\mathbf{H}_\text{TR}) + \frac{\lambda_{\min}^2(\mathbf{H}_\text{TX}) \lambda_{\min}^2(\mathbf{H}_\text{XR})}{\lambda_{\max}^2(\mathbf{H}_\text{TX}) + \lambda_{\min}^2(\mathbf{H}_\text{XR}) + N_\text{T}}. && (2.37)
\end{aligned}$$

2.5.1.2 Single Antenna Selection Diversity Analysis

The post-processing SNR for the received data stream is obtained after some manipulations as

$$\text{SNR}_{\text{(RC)},1}(1, p, 1, q) = \gamma_\text{T} \|\mathbf{h}_{\text{TR},p}\|^2 + \frac{\gamma_\text{T} \gamma_\text{X} \|\mathbf{h}_{\text{TX},p}\|^2 \|\mathbf{h}_{\text{XR},p}\|^2}{\gamma_\text{T} \|\mathbf{h}_{\text{TX},p}\|^2 + \gamma_\text{X} \|\mathbf{h}_{\text{XR},p}\|^2 + 1}. \tag{2.38}$$

The expression in (2.38) is equivalent to (14) in [154], which makes it possible to select the single best antennas at the transmitter and at the relay independently. I denote $q^* \triangleq \arg\max_{1 \leq q \leq N_\text{X}} \|\mathbf{h}_{\text{XR},p}\|$, $p^* \triangleq \arg\max_{1 \leq p \leq N_\text{T}} \text{SNR}_{\text{(RC)},1}(1, p, 1, q^*)$, and $\text{SNR}_{\text{(RC)}}^{(s)} \triangleq \text{SNR}_{\text{(RC)},1}(1, p^*, 1, q^*)$. The following upper bound on $\text{SNR}_{\text{(RC)}}^{(s)}$ is obtained based on (2.38), (2.22), and the inequality $\max_{1 \leq p \leq N_\text{T}} \|\mathbf{h}_{\text{TR},p}\|^2 \leq$

$\sigma^2_{\max}(\mathbf{H}_{\text{TX}})$ [83]

$$\begin{aligned} \text{SNR}^{(s)}_{(\text{RC})} &\leq \lambda^2_{\max}(\mathbf{H}_{\text{TR}}) + \frac{\lambda^2_{\max}(\mathbf{H}_{\text{TX}})\lambda^2_{\max}(\mathbf{H}_{\text{XR}})}{\lambda^2_{\max}(\mathbf{H}_{\text{TX}}) + \lambda^2_{\min}(\mathbf{H}_{\text{XR}}) + 1} \quad (2.39) \\ &\triangleq \lambda^2_{\max}(\mathcal{H}^{(m)}_{(\text{RC})}). \quad (2.40) \end{aligned}$$

In (2.40), the maximum SNR-scaled singular value of $\mathcal{H}^{(m)}_{(\text{RC})}$ is defined as the upper bound in (2.39). The upper bound in (2.39) shows the relationship between the SNR performance of full selection diversity and the eigenmodes of the original constituent channels. Note that $\lambda^2_{\max}(\mathcal{H}^{(m)}_{(\text{RC})}) = \lambda^2_{\max}(\mathbf{H}_{\text{TR}}) + \lambda^2_{\max}(\mathcal{H}^{(m)}_{(\text{TH})})$. This means that the upper bound on $\text{SNR}^{(s)}_{(\text{RC})}$ is the sum of the upper bound on $\text{SNR}^{(s)}_{(\text{TH})}$ and the upper bound on $\text{SNR}^{(s)}_{(\text{TR})}$, where $\text{SNR}^{(s)}_{(\text{TR})}$ is the maximum post-processing SNR of the single-stream antenna selection diversity on the direct channel. Interestingly, the upper bound given in (2.39) is the same as the counterpart in [157], where the receiver uses the nonlinear ML receive filters (the system model in this chapter considers the linear ZF receive filters).

2.5.2 Switching Between Spatial Multiplexing and Selection Diversity

Note that the relay channel dualmode selection criteria are natural extensions of the two-hop dualmode selection criteria. In this section, I provide the list of selection criteria just for the completeness.

Relay Channel Selection Criterion 1 - VSER-based Selection: Choose spatial multiplexing if

$$P^{(m)}_{e,(\text{RC})} < P^{(s)}_{e,(\text{RC})};$$

otherwise, choose selection diversity transmission with the best antennas at the transmitter and relay.

Relay Channel Selection Criterion 2 - Post-Processing SNR-based Selection: Choose spatial multiplexing if

$$d_{\min}^2(N_{\mathrm{T}}, R)\mathrm{SNR}_{(\mathrm{RC})}^{(m)} < d_{\min}^2(1, R)\mathrm{SNR}_{(\mathrm{RC})}^{(s)};$$

otherwise, choose selection diversity transmission with the best antennas at the transmitter and relay.

With the minimum and maximum singular values of the relay channel defined in (2.36) and (2.40), I define the relay condition number as

$$\kappa_{(\mathrm{RC})} \triangleq \frac{\lambda_{\max}\left(\mathcal{H}_{(\mathrm{RC})}^{(m)}\right)}{\lambda_{\min}\left(\mathcal{H}_{(\mathrm{RC})}^{(s)}\right)}. \tag{2.41}$$

It follows from the lower bound of $\lambda_{\min}^2(\mathcal{H}_{(\mathrm{RC})}^{(m)})$ in (2.37) that obtain

$$\kappa_{(\mathrm{RC})}^2 \leq \frac{\lambda_{\max}^2(\mathbf{H}_{\mathrm{TR}}) + \frac{\lambda_{\max}^2(\mathbf{H}_{\mathrm{TX}})\lambda_{\max}^2(\mathbf{H}_{\mathrm{XR}})}{\lambda_{\max}^2(\mathbf{H}_{\mathrm{TX}})+\lambda_{\min}^2(\mathbf{H}_{\mathrm{XR}})+1}}{\lambda_{\min}^2(\mathbf{H}_{\mathrm{TR}}) + \frac{\lambda_{\min}^2(\mathbf{H}_{\mathrm{TX}})\lambda_{\min}^2(\mathbf{H}_{\mathrm{XR}})}{\lambda_{\max}^2(\mathbf{H}_{\mathrm{TX}})+\lambda_{\min}^2(\mathbf{H}_{\mathrm{XR}})+N_{\mathrm{T}}}}. \tag{2.42}$$

The right-hand side in (2.42) is exactly the aggregate condition number proposed in [154], which is derived for the relay channel with the uniform power allocation at the relay and the nonlinear ML receive filters at the receiver. Note that the relay condition number is proposed for the relay channel with the MSE-minimization power allocation at the relay and the linear ZF receive filters at the receiver. Although not explicitly defined in [154], the lower bound of $\lambda_{\min}^2(\mathcal{H}_{(\mathrm{RC})}^{(m)})$ in (2.37) can be treated as the minimum singular value of the specific relay channel considered in [154]. Moreover, the MSE-minimization relay filter in this chapter can compensate for the effect of noise enhancement resulted from the use of the ZF receive filters. Thus, it leads to a higher minimum singular value of the corresponding relay channel and hence a lower effective condition number. This means that the relay channel considered in this chapter is more appropriate for spatial multiplexing transmission.

By substituting (2.34), (2.40), and (2.41) into *Relay Channel Selection Criterion 2*, I obtain the following relay channel multimode antenna selection criterion.

Relay Channel Selection Criterion 3 - Condition Number-based Selection: Choose spatial multiplexing if

$$\kappa^2_{(\mathrm{RC})} < \frac{1}{\sqrt{N_\mathrm{T}}} \frac{d_{\min}(N_\mathrm{T}, R)}{d_{\min}(1, R)};$$

otherwise, choose selection diversity transmission with the best antennas at the transmitter and relay.

2.5.3 Multimode Selection with the Direct Link

The first two multimode selection criteria when the direct link is considered have the same form as those for the two-hop channel, except that the expression of the post-processing SNR is determined based on (2.11) instead of (2.10). Therefore, I just list the criteria here for the sake of completeness.

Relay Channel Selection Criterion 4 - VSER-based Selection: Choose

ω^* that solves

$$\arg\min_{\omega\in\Omega} 1-\left[1-N_eQ\left(\sqrt{\mathrm{SNR}_{(\mathrm{RC})}(\omega)\frac{d^2_{\min}(\omega)}{2}}\right)\right]^{m_\mathrm{T}}. \tag{2.43}$$

Relay Channel Selection Criterion 5 - SNR-based Selection: Choose ω^* that solves

$$\arg\min_{\omega\in\Omega} d^2_{\min}(\omega)\mathrm{SNR}_{(\mathrm{RC})}(\omega). \tag{2.44}$$

The main focus in this section is on the eigenmode-based multimode selection criterion. First, (2.30) is rewritten as $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})}) = \lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{TR}) + \lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{TH})})$. Second, it follows from Theorem 3 in [83] that $\lambda_{m_\mathrm{T}}(\mathbf{H}_\mathrm{TR}) \geq \lambda_{m_\mathrm{T}}(\mathcal{H}_\mathrm{TR})$. Lastly, an approximation of $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})})$ is obtained by using the approximation of $\lambda_{m_\mathrm{T}}(\mathcal{H}_\mathrm{TR})$ and the approximation of $\lambda_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{TH})})$ in (2.28) into (2.30). It is given by

$$\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})}) \approx \lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{TR}) + \frac{\lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{TX})\lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{XR})}{\frac{1}{m_\mathrm{T}}\sum_{i=1}^{m_\mathrm{T}}\lambda^2_i(\mathbf{H}_\mathrm{TX}) + \lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{XR}) + m_\mathrm{T}}. \tag{2.45}$$

Note that the approximation of $\lambda^2_{m_\mathrm{T}}(\mathcal{H}_{(\mathrm{RC})})$ in (2.45) depends only on m_T and the original constituent channels. Thus, the optimal number of streams can be determined separately from the other parameters. I propose the following eigenmode-based multimode antenna selection criterion.

Relay Channel Selection Criterion 6 - Eigenmode-based Selection Criterion: The selection procedure consists of three steps:

Step 1: Find m^*_T that solves

$$\arg\max_{m_\mathrm{T}\in\mathcal{M}_\mathrm{T}} \frac{d^2_{\min}(m_\mathrm{T},R)}{m_\mathrm{T}}\left[\lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{TR}) + \frac{\lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{TX})\lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{XR})}{\frac{1}{m_\mathrm{T}}\sum_{i=1}^{m_\mathrm{T}}\lambda^2_i(\mathbf{H}_\mathrm{TX}) + \lambda^2_{m_\mathrm{T}}(\mathbf{H}_\mathrm{XR}) + m_\mathrm{T}}\right]$$

Step 2: Find (m_{X}^*, q^*) that solve

$$\underset{m_{\mathrm{T}}^* \leq m_{\mathrm{X}} \leq M_{\mathrm{X}}, 1 \leq q \leq \binom{N_{\mathrm{X}}}{m_{\mathrm{X}}}}{\arg\max} \lambda_{m_{\mathrm{T}}^*}(\mathbf{H}_{\mathrm{XR}} \mathbf{W}_{\mathrm{X},(m_{\mathrm{X}},q)})$$

Step 3: Choose p^* that solves

$$\underset{1 \leq p \leq \binom{N_{\mathrm{T}}}{m_{\mathrm{T}}^*}}{\arg\max} \left[\lambda_{m_{\mathrm{T}}^*}^2(\mathbf{H}_{\mathrm{TR}}) + \frac{\lambda_{m_{\mathrm{T}}^*}^2(\mathbf{H}_{\mathrm{TX}}) \lambda_{m_{\mathrm{T}}^*}^2(\mathbf{H}_{\mathrm{XR}})}{\frac{1}{m_{\mathrm{T}}^*} \sum_{i=1}^{m_{\mathrm{T}}^*} \lambda_i^2(\mathbf{H}_{\mathrm{TX}}) + \lambda_{m_{\mathrm{T}}^*}^2(\mathbf{H}_{\mathrm{XR}}) + m_{\mathrm{T}}^*} \right].$$

Compared to the other multimode algorithms, the eigenmode-based algorithm has lower computational complexity, which comes at the price of accuracy as several approximations are used in the derivation. The simulations in Section 2.6 show that this algorithm still provides a large array gain over all dualmode algorithms. Therefore, when computational complexity is important, this multimode selection criterion is the best candidate. More importantly, this algorithm shows how the quality of the original constituent channels affects multimode transmission.

2.6 Simulations

This section provides Monte Carlo simulations to evaluate the VSER performance of the proposed antenna selection criteria. Section 2.6.1 describes and discusses the link-level simulations for the system model considered in the chapter so far. Section 2.6.2 presents the multi-cell simulations with more realistic channel models and a hexagonal cellular network.

2.6.1 Link-Level Simulations

I assume all constituent channels are subject to frequency-flat block fading with an i.i.d. Rayleigh fading model. Also, SNR_{TX}, SNR_{XR} and SNR_{TR} denote the mean SNR values of the corresponding channels. Due to space constraints, this section considers only the systems where each node has four antennas and the receiver uses a linear ZF receive filter. The overall data rate is fixed at $R = 8$ bits and symbols are drawn from QAM constellations with unit power. The dualmode system supports only two cases: i) $m_{\text{T}} = m_{\text{X}} = 1$ with 256-QAM and ii) $m_{\text{T}} = m_{\text{X}} = 4$ with 4-QAM. The multimode system can support $m_{\text{T}} \in \{1, 2, 4\}$ with 256-QAM, 16-QAM, and 4-QAM, respectively. The VSER results of the proposed antenna selection algorithms are compared with beamforming approaches [109]. Grassmannian codebooks from [129] are used for limited feedback beamforming. It follows the same proof in [157] that the single best antenna selection achieves the full diversity order of both the two-hop channel (which is $N_{\text{X}} \min\{N_{\text{T}}, N_{\text{R}}\}$) and the relay channel (which is $N_{\text{T}}N_{\text{R}} + N_{\text{X}} \min\{N_{\text{T}}, N_{\text{R}}\}$).

Experiment 1-Antenna Selection for the Two-Hop Channel: Fig. 2.2 shows the results for the two-hop channel with fixed SNR_{TX} (at 20 dB) and varying SNR_{XR}. There is little difference between the first two dualmode algorithms, which have a gain of about 2.5 dB gain at 10^{-1} over the third dual algorithm. The simulation shows that although suboptimal, all dualmode algorithms achieve the full diversity order as the single best antenna selection. In addition, the dualmode algorithms provide array gains over the single best

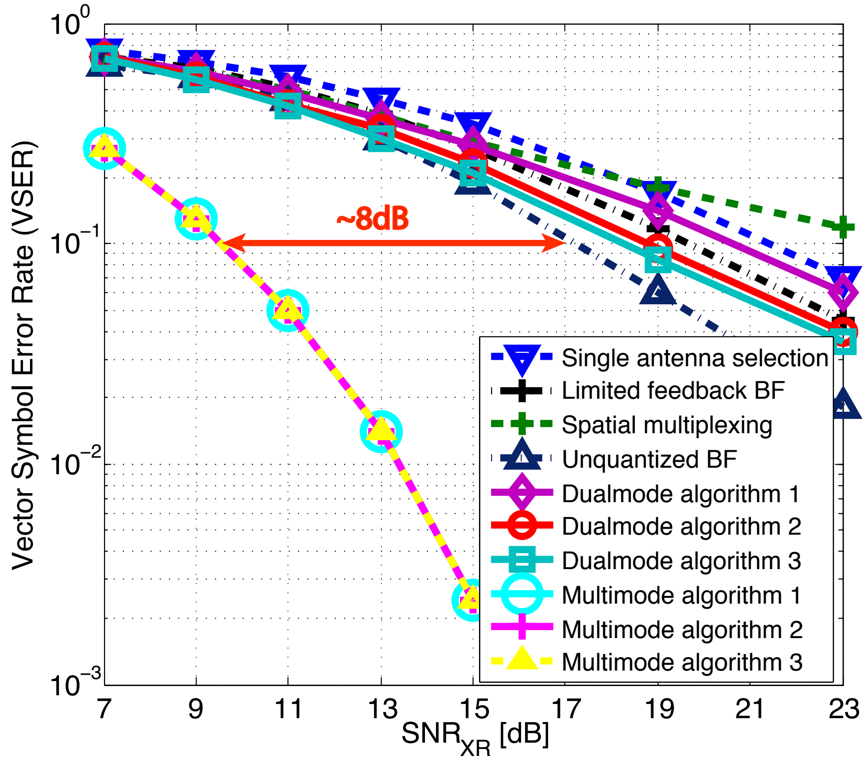

Figure 2.2: VSER performance of dualmode and multimode algorithms for the two-hop channel. The proposed algorithms achieve full diversity order. In addition, the proposed multimode algorithms provide considerable array gains over single-stream transmission strategies.

antenna selection (approximately 3 dB at 10^{-1} for the first two algorithms with only one more feedback bit). These results mean that the dualmode algorithms make a quite exact selection of the better mode. The first two dualmode algorithms also perform better than limited feedback Grassmannian beamforming (0.5 dB gain at 10^{-1}) and perform very closely to the optimal full channel knowledge (i.e., unquantized) beamforming in the low SNR regime (since in this regime full spatial multiplexing provides a lower VSER than full selection diversity). Finally, all the two-hop multimode antenna selection criteria give nearly the same VSER performance. A similar observation has been made in the plot where $\mathrm{SNR}_{\mathrm{XR}}$ is fixed and $\mathrm{SNR}_{\mathrm{TX}}$ is varied. Such observations are important since they allow the use of the lowest-complexity multimode algorithm, i.e., the eigenmode-based multimode selection criterion. Notably, the multimode algorithms provide a large array gain (around 8 dB at 10^{-1}) over the optimal unquantized beamforming, the best single-stream transmission technique in terms of VSER performance. This is because by supporting a varying number of data streams, multimode transmission provides better adaptation of transmitted signals to the channels than single-stream transmissions.

Experiment 2-Antenna Selection for the Relay Channel: To focus on the impact of the direct link on the VSER performance of relay channel, I fix $\mathrm{SNR}_{\mathrm{TX}} = 20$ dB and $\mathrm{SNR}_{\mathrm{XR}} = 10$ dB and produce the curves of VSER values against $\mathrm{SNR}_{\mathrm{TR}}$. With the values of $\mathrm{SNR}_{\mathrm{TX}}$ and $\mathrm{SNR}_{\mathrm{XR}}$, Fig. 2.2 informs that full spatial-multiplexing is preferred to full selection diversity. Moreover, if only the direct channel is used, full spatial multiplexing on average performs better than full selection diversity [83] at low $\mathrm{SNR}_{\mathrm{TR}}$. As expected in Fig. 2.3, in the low $\mathrm{SNR}_{\mathrm{TR}}$ regime, full spatial multiplexing gives a lower average VSER than full selection diversity, and even than the limited feedback Grassmannian beamforming. Notice that the proposed dualmode and multimode algorithms, which contain full selection diversity as a special mode, can obtain the full diversity gain of the relay channel. The three multimode algorithms no longer have the same VSER performance curves. The eigenmode-based multimode

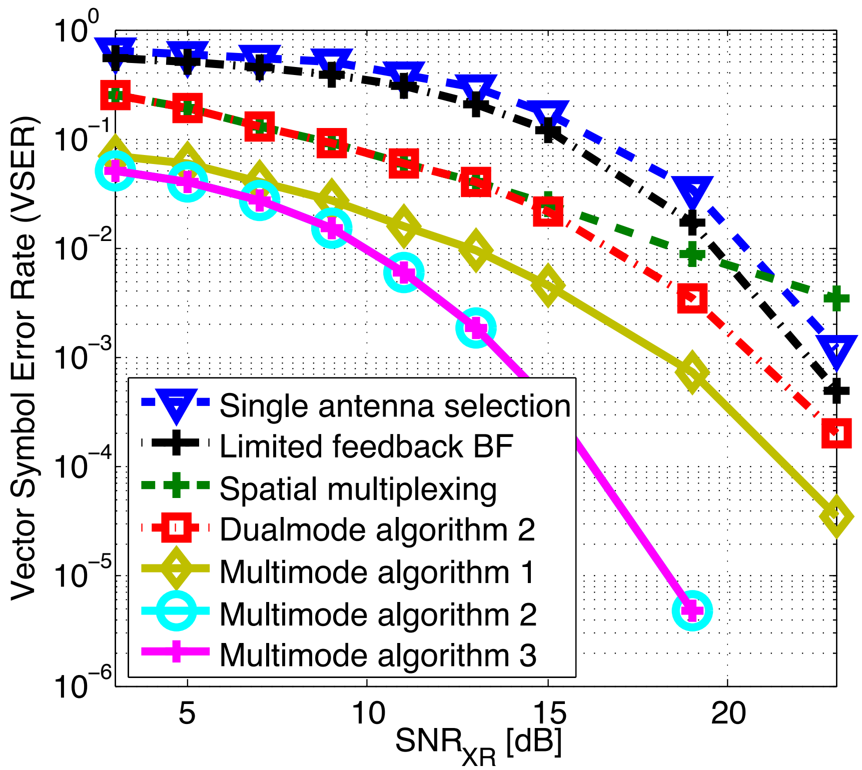

Figure 2.3: Comparison of VSER performance of several transmission strategies for the relay channel. The proposed multimode algorithms obtain full diversity order and provide large array gains over single-stream transmission strategies.

algorithm has lower computational complexity at the cost of an array loss of 3.5 dB in relative comparison with the other multimode algorithms. The eigenmode-based multimode algorithm, however, seems to provide the best tradeoff between computational complexity and VSER performance among the proposed antenna selection criteria. In particular, it provides an array gain of 4.5 dB at 10^{-2} over the SNR-based dualmode algorithm (at the price of higher computational complexity).

Experiment 3-Impact of Feedback Delays: In the experiment, I investigate the impact of feedback delays on the VSER performance of the multimode algorithms in a downlink time-varying relay channel (where I assume the transmitter-relay channel is stable over time). Specifically, the receiver is assumed to move at the speed of $v = 30$ km/h, the operating frequency is $f = 2$ GHz and the sampling period is $T_s = 1$ ms. For the transmitter-receiver and relay-receiver channels, the normalized channel at time n is denoted as $\mathbf{H}[n]$. The complex elements of the normalized channels are drawn from an arbitrary i.i.d. distribution with zero mean and unit variance. Let α be the temporal correlation computed by the zero-order Bessel function $\alpha = J_0(2\pi T_s f v/c)$, $c = 3 \times 10^8$ m/s be the speed of light, and $\mathbf{H}_w$ be a matrix of the relevant size with i.i.d. complex Gaussian elements of zero mean and unit variance. I assume that $\mathbf{H}[n]$ is generated based on $\mathbf{H}[n-1]$ according to the first order autoregression fading model (Gauss-Markov fading model) as [15]

$$\mathbf{H}[n] = \alpha \mathbf{H}[n-1] + \sqrt{1-\alpha^2}\mathbf{H}_w. \tag{2.46}$$

The feedback delay is characterized by D. This means that a mode, which was

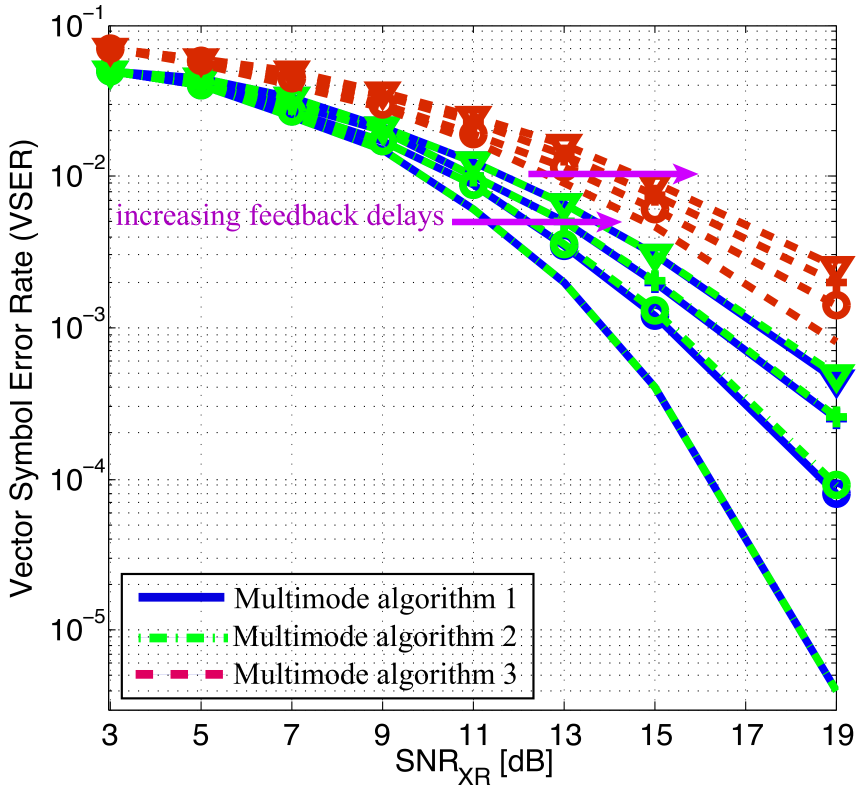

Figure 2.4: Impact of feedback delays on the VSER performance of the proposed algorithms. The mismatch between the channels based on which a mode is designed and the channels for which the mode is applied degrades the performance. The curves without markers are for the case without feedback delays. The circle, cross, and triangle markers denote the curves corresponding to the feedback delays (D) of 10, 20, and 30, respectively.

designed based on the channels at time n, is actually applied to the channels at time $(n+D)$. This experiment considers three different values of D, including 10, 20, and 30. Fig. 2.4 shows that VSER performance of the multimode algorithms degrades with the increase of feedback delay. Note that the losses at 10^{-2} are less than 2 dB.

2.6.2 Multi-Cell Simulations

Like much prior work, this chapter so far did not consider interference. In cellular networks, for example, in addition to path loss and heavy shadowing, the out-of-cell interference degrades significantly the direct link between a base station and its served mobile station located near the boundary of the base station's cell area. In this section, the proposed two-hop multimode algorithms are evaluated under the multi-cell simulations to show that the use of relays can improve the reliability of downlink transmission to cell-edge users. I adopt the system model for the simulations from [159], which considers the realistic channel models and a layer of interfering cells wrapped around the three main cells. The cells have hexagonal shapes. The differences are that I consider multiple-antenna nodes, where the parameters for each MIMO link are the same as in Section 2.6.1, and that the distance from the RS to the cell corner is one-third of the cell radius. This experiment also investigates the one-hop algorithm in [83], that sends two different vector symbols in two stages, each carrying four bits.

Remark 2.6.1. Note that the results work for any topology of base stations, not just the hexagonal grid model of base station locations as considered in this Subsection. In real cellular networks, due to practical deployment constraints, the actual locations of base stations are more random [220]. Recent results propose a new model where the base stations are randomly located as a homogeneous Poisson point process [13]. This approach provides more tractable analysis of interference and hence insights on cellular network performance. Future work may use this framework to investigate the impact of interference on my results.

Fig. 2.5 shows the VSER as a function of the MS distance from the cell corner for frequency reuse factor of six for various transmission strategies. As the MS moves far away from the RS, the RS-MS channel degrades due to increased path loss, thus the VSER values of two-hop transmission strategies increase. Moreover, the two-hop multimode algorithm provides much lower VSER than the two-hop full spatial multiplexing and full selection diversity. This means that to a certain degree the proposed algorithm can adapt well to the interference-limited environment. The simulation results also show that the two-hop multimode algorithm outperforms the one-hop multimode transmission when the MS distance from the cell corner is less than two-third of the cell radius. For example, the two-hop algorithm can obtain an average VSER of 10^{-1} when the MS is located at the cell corner; while the one-hop algorithm requires the MS position at half of the cell radius from the BS to achieve the same value of VSER. Nevertheless, the one-hop algorithm performs better when the MS is close to the BS due to decreased path-loss, shadowing,

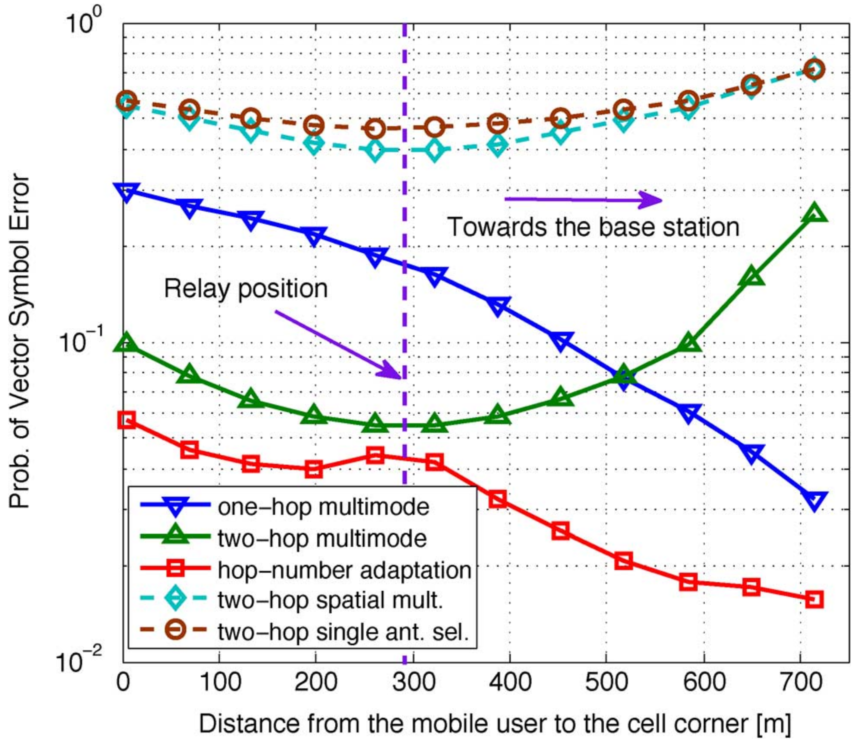

Figure 2.5: VSER performance of one-hop and two-hop multimode algorithms in a cellular network. The total number of bits transmitted in two stages is eight bits. The two-hop algorithm outperforms for cell-edge users while the one-hop algorithm is better for users located close to the base station.

and out-of-cell interference. In eigenmode-based multimode algorithms, the performance metric is characterized by the product of the square of minimum constellation distance and the minimum SNR-scaled singular value of the end-to-end channel. I also develop a *hop-number adaptation* algorithm which selects adaptively the number of hops by comparing comparing the values for the selected one-hop and two-hop modes. In Fig. 2.5, the hop-number adaptation is shown to outperform both the one-hop and two-hop multimode algorithms in all ranges of MS locations.

Chapter 3

Cooperative Algorithms for MIMO Amplify-and-Forward Relay Networks

This chapter presents cooperative algorithms for jointly configuring the transmitters, relays, and receivers in the MIMO AF relay interference channel. Section 3.1 presents the motivations, reviews prior work, and introduces my contributions. Section 3.2 describes the system model. Section 3.3 and Section 3.4 present the proposed algorithms in detail. Section 3.5 discusses their properties while Section 3.6 numerically evaluates their achievable end-to-end sum-rates and multiplexing gains.

3.1 Introduction

The relay interference channel models a network where a stage of intermediate nodes, called relays, help multiple transmitters communicate with their receivers using shared radio frequency resources [30, 190, 203, 204]. Relay communication is considered a viable solution for coverage extension and capacity enhancement [5, 98]. Recent results show that the single-hop interference channel (without relays) may not be interference limited under certain conditions [27, 73, 100, 152, 166, 225]. Although these single-hop results can be applied separately for the transmitter-relay hop and for the relay-receiver hop, higher end-to-end sum-rates can be achieved if the relays are configured jointly [66, 103]. Obtaining the most from the relay interference channel requires advanced interference management strategies that jointly configure the transmitters, relays, and receivers.

Multiplexing gain is an important performance metric of interference networks. The multiplexing gain of a network, also known as the total number of degrees of freedom, is a first-order approximation of its sum capacity at high SNR [91]. Interference alignment is a technique that maximizes the multiplexing gain of the single-hop interference channel [100]. The idea is to arrange the transmitted signals such that interference is constrained within only a portion of the signal space observed by each receiver. This leaves the remaining portion for interference-free detection of desired signals [27]. The maximum multiplexing gain achievable through interference alignment, however, depends on the characteristics of the interference channel. For a symmetric MIMO single-hop interference channel with constant channel coefficients, the maximum multiplexing gain is upper-bounded by the total number of antennas at the transmitter and receiver of each pair, independently of the number of pairs in the network [166, 225]. Note that the bound is tight in certain cases and corresponds to the total available space dimensions at each pair. Increasing the number of space dimensions in the network, using for example relays, is one way to improve the maximum achievable multiplexing gain.

I consider the half-duplex multiple-antenna AF relay interference channel. Several interference management strategies designed specifically for the one-way AF relay interference channel have been proposed in [7, 28, 33, 38, 41, 70, 74, 103, 120, 140, 145, 147, 163]. In [70], relays were shown to reduce the number of independent channel extensions needed to align interference at the receivers, despite their inability to improve the multiplexing gains of the single antenna fully connected interference channel for time-varying or frequency-selective channel coefficients [28].

Prior work often considered networks operating in special circumstances. In [7, 41, 103, 140, 147, 163], it was assumed that there are enough antennas at the relays to both cancel received interference, and null interference caused to other users when retransmitting, allowing the achievable multiplexing gain to scale linearly with the number of users. In [33, 74, 120], a small network with up to three transmitter-receiver pairs was considered to derive closed-form solutions. In [38], design problems with different objective functions including sum power minimization and minimum SINR maximization were considered. In this chapter, I consider a general system model in the sense that I assume no special constraints on the number of relays or the number of antennas at a relay. The closest AF relay model to mine was assumed in [145], but it considered single-antenna transmitters and receivers. Furthermore, the authors in [145] assumed no crosslinks from relays to receivers, resulting in an oversimplified design problem.

In this chapter, I develop three cooperative interference management algorithms for the MIMO AF relay interference channel with constant channel coefficients. I assume that global CSI is available at a central processing unit for jointly designing the transmitters, relays, and receivers. The first algorithm is an extension of the total leakage minimization interference alignment algorithm for the single-hop interference channel in [69, 70, 156, 158]. Note that the leakage signals in AF relay interference networks consist of interference signals and enhanced relay noise. The first algorithm is useful for gaining insights into the interference alignment feasibility of the MIMO AF relay interference channel. The second algorithm aims at directly finding the stationary points of the end-to-end sum-rate maximization problem with equality constraints on the transmit power; while the third algorithm deals specifically with inequality constraints on the transmit powers. Similar to the single-hop results in [165, 187], my second and third algorithms are inspired by a connection between achievable rates and MSE values. The key observation is that there exist matrix-weighted sum-MSE minimization problems that have the same stationary points as the sum-rate maximization problems provided the matrix weights are chosen appropriately. The formulated matrix-weighted sum-MSE minimization problems are non-convex and NP-hard; finding their globally optimal solutions is challenging. I propose to use alternating minimization where in each iteration, I fix all but one variable and focus on determining the remaining variable. It can be proved that the proposed MSE-based algorithms always converge to the stationary points of the associated optimization problems. Note that the power constraints at the relays depend on both the precoders at the transmitters and the processing matrices at the relays, thereby compounding the design problems. It is thus not straightforward to extend the methods used for the single-hop design problems to solve the two-hop problems.

I use Monte Carlo simulations to evaluate the average end-to-end sum-rates and multiplexing gains achievable through the proposed algorithms. I first verify the convergence of the proposed algorithms. They work as expected thanks to their ability of finding the global optimum of the corresponding single-variable optimization problem at each iteration. I then show that as the number of iterations increases, the effect of interference, which started off as dominant, decreases, until it is canceled. Enhanced noise from the relays subsequently becomes the dominant factor. This confirms the importance of taking into account enhanced relay noise when designing relay-aided interference alignment algorithms. I further show that the total leakage minimization algorithm achieves lower average end-to-end sum-rates when compared to the MSE-based algorithms at low to medium SNR values. This results from neglecting the desired signal and noise powers at the receivers. The MSE-based algorithms, nevertheless, result in unfairness, i.e., some users have much smaller rates than the others. Thus, the MSE-based algorithms achieve lower average end-to-end sum-rates and multiplexing gains than the total leakage minimization algorithm at high SNR. One reason for this is that the MSE-based algorithms are not guaranteed to find the global optima of the end-to-end sum-rate maximization problems. I further show that for a fixed number of antennas at the transmitters and receivers, AF relays can provide larger

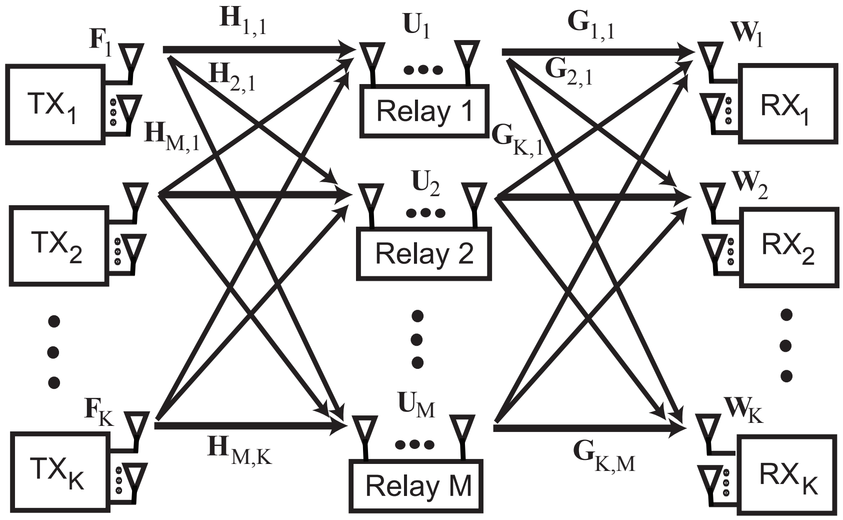

Figure 3.1: A relay interference channel where M half-duplex AF relays aid the one-way communication of K transmitter-receiver pairs.

end-to-end multiplexing gains than DF relays or direct transmission, despite the half-duplex loss. I finally show that AF relays provide larger average achievable end-to-end sum-rates than DF relays. The proposed algorithms in general provide higher achievable end-to-end sum-rates than the baseline AF relaying strategies that do not align interference at the receivers.

3.2 System Model

In this chapter, I consider a relay interference channel where M half-duplex AF relays aid the one-way communication between K pairs of transmitters and receivers, as illustrated in Fig. 3.1. Each transmitter has data for only one receiver and each receiver is served by only one transmitter. Each pair is assigned a unique index $k \in \mathcal{K} \triangleq \{1, \cdots, K\}$. Transmitter k has $N_{\mathrm{T},k}$

antennas while receiver k has $N_{\mathrm{R},k}$ antennas for $k \in \mathcal{K}$. Similarly, each relay is assigned a unique index $m \in \mathcal{M} \triangleq \{1, \cdots, M\}$. Relay m has $N_{\mathrm{X},m}$ antennas for $m \in \mathcal{M}$. Since half-duplex relays cannot transmit and receive at the same time by assumption, the transmission procedure consists of two stages. In the first stage, the transmitters send data to the relays. In the second stage, the relays apply linear processing to their observed signals and forward them to the receivers. I assume the direct channels between the transmitters and receivers are ignored by the second-stage receivers.

I assume the relay interference channel is fully-connected on each hop. Specifically, each relay receives non-negligible signals from all the transmitters and each receiver observes non-negligible signals from all the relays. This assumption is good for modeling certain scenarios in cellular networks. For example, relays are used to provide indoor coverage for a large building. Due to penetration loss and shadowing, the direct channels from base stations to the mobile users located in the building are weak and negligible. Relays are deployed at good geographical locations, e.g., with high altitudes, around the building to aid the base stations communicate with the users. Relays on different sides of the building receive aid different base stations to their associated users. On the backhaul links from the base stations to the relays, due to their high altitudes, each relay observes non-negligible signals from the base stations. Similarly, on the access links from the relays to the users, each user receives non-negligible signals from the relays.

Let $\mathbf{H}_{m,k} \in \mathbb{C}^{N_{\mathrm{X},m} \times N_{\mathrm{T},k}}$ be the matrix channel from transmitter k to relay m and $\mathbf{G}_{k,m} \in \mathbb{C}^{N_{\mathrm{R},k} \times N_{\mathrm{X},m}}$ be the matrix channel from relay m to receiver k for $k \in \mathcal{K}$ and $m \in \mathcal{M}$. I assume that perfect and instantaneous knowledge of $\mathbf{H}_{m,k}$ and $\mathbf{G}_{k,m}$ for $k \in \mathcal{K}$ and $m \in \mathcal{M}$ is available at a central processing unit. Note that methods for obtaining the global CSI in the single-hop interference channel can be used for the two hops of the relay systems. Specifically, each relay measures the channels from the transmitters to itself and sends to the central unit. Similarly, each receiver measures the channels from the relays to itself and sends to the central unit. Papers that make such assumptions like [28, 41, 92, 143, 161] are valuable because they give a benchmark for developing algorithms that relax these assumptions.

I assume Gaussian signaling is used in the system although it may not be optimal. Specifically, the transmitters, including uncoordinated interferers, use independent Gaussian codebooks for transmission. This allows the receivers to treat interference as additive Gaussian noise while decoding their desired signals. Let $\mathbf{n}_{\mathrm{X},m}$ be the spatially white, additive Gaussian noise at relay m with covariance $\mathbb{E}(\mathbf{n}_{\mathrm{X},m}\mathbf{n}_{\mathrm{X},m}^{*}) = \sigma_{\mathrm{X},m}^{2}\mathbf{I}_{N_{\mathrm{X},m}}$ for $m \in \mathcal{M}$. In real systems, however, wireless transceivers are also affected by extra sources of interference, of which the statistical distribution is nonGaussian [46, 47]. For example, interference may be generated from atmospheric noise and electrical discharge. Interference may be generated from clocks and buses in the computational platform where the wireless transceivers are deployed. I assume that nonGaussian interference and noise are out of the scope of this chapter.

Let $\mathbf{s}_k \in \mathbb{C}^{d_k \times 1}$ be the transmit symbol vector at transmitter k, where

$d_k \leq \min\{N_{\mathrm{T},k}, N_{\mathrm{R},k}\}$ is the number of data streams from transmitter k to receiver k for $k \in \mathcal{K}$. The transmit symbols are independent and identically distributed (i.i.d.) such that $\mathbb{E}(\mathbf{s}_k\mathbf{s}_k^*) = \mathbf{I}_{d_k}$. Transmitter k uses a linear transmit precoder $\mathbf{F}_k \in \mathbb{C}^{N_{\mathrm{T},k} \times d_k}$ to map $\mathbf{s}_k$ to its antennas. The transmit power at transmitter k is $p_{\mathrm{T},k} = \mathrm{tr}(\mathbf{F}_k^*\mathbf{F}_k)$. Let $p_{\mathrm{T},k}^{\max}$ be the maximum transmit power. With perfect synchronization, relay m observes the following signal

$$\mathbf{y}_{\mathrm{X},m} = \sum_{k=1}^{K} \underbrace{\mathbf{H}_{m,k}\mathbf{F}_k}_{\boldsymbol{\mathcal{H}}_{m,k}} \mathbf{s}_k + \mathbf{n}_{\mathrm{X},m}. \tag{3.1}$$

Let $\mathbf{U}_m \in \mathbb{C}^{N_{\mathrm{X},m} \times N_{\mathrm{X},m}}$ be the processing matrix at relay m. The transmit signal at relay m is given by

$$\begin{aligned} \mathbf{x}_{\mathrm{X},m} &= \mathbf{U}_m \mathbf{y}_{\mathrm{X},m} \\ &= \sum_{k=1}^{K} \mathbf{U}_m \boldsymbol{\mathcal{H}}_{m,k}\mathbf{s}_k + \mathbf{U}_m \mathbf{n}_{\mathrm{X},m}. \end{aligned} \tag{3.2}$$

The transmit power at relay m is thus computed as follows

$$p_{\mathrm{X},m} = \sum_{k=1}^{K} \mathrm{tr}(\mathbf{U}_m \boldsymbol{\mathcal{H}}_{m,k} \boldsymbol{\mathcal{H}}_{m,k}^* \mathbf{U}_m^*) + \sigma_{\mathrm{X},m}^2 \,\mathrm{tr}(\mathbf{U}_m \mathbf{U}_m^*). \tag{3.3}$$

There are two possible types of power constraints at the relays: i) a set of individual power constraints at the relays, and ii) a sum power constraint at all the relays. Per-relay power constraints are often considered in the cellular system literature [162, 168]. A sum power constraint is however considered in the ad hoc network literature to extend the lifetime of battery-powered relays [57, 102, 121]. Let $p_{\mathrm{X},m}^{\max}$ be the maximum transmit power at relay m

and $p_{\mathrm{X}}^{\max}$ be the maximum sum transmit power at all the relays. When power control is considered, the per-relay power constraints are

$$p_{\mathrm{X},m} \;\; \leq \;\; p_{\mathrm{X},m}^{\max}, \forall m \in \mathcal{M}, \tag{3.4}$$

whereas the sum power constraint on the relays is

$$\sum_{m=1}^{M} p_{\mathrm{X},m} \;\; \leq \;\; p_{\mathrm{X}}^{\max}. \tag{3.5}$$

Without power control, the inequalities in (3.4) and (3.5) are replaced by equalities. The following sections focus on the sum power constraint at the relays. The reason for this assumption and the applicability of per-relay power constraints are discussed in Section 3.5.

Let $\mathbf{n}_{\mathrm{R},k}$ be the spatially white, additive Gaussian noise at receiver k with covariance $\mathbb{E}(\mathbf{n}_{\mathrm{R},k}\mathbf{n}_{\mathrm{R},k}^{*}) = \sigma_{\mathrm{R},k}^{2}\mathbf{I}_{N_{\mathrm{R},k}}$ and $\mathcal{G}_{k,m} = \mathbf{G}_{k,m}\mathbf{U}_{m}$. Receiver k observes the following signal

$$\begin{aligned} \mathbf{y}_k &= \sum_{m=1}^{M} \mathbf{G}_{k,m}\mathbf{x}_{\mathrm{X},m} + \mathbf{n}_{\mathrm{R},k} \\ &= \sum_{q=1}^{K} \underbrace{\sum_{m=1}^{M} \mathcal{G}_{k,m}\mathcal{H}_{m,q}}_{\mathcal{T}_{k,q}} \mathbf{s}_q + \sum_{m=1}^{M} \mathcal{G}_{k,m}\mathbf{n}_{\mathrm{X},m} + \mathbf{n}_{\mathrm{R},k}, \end{aligned}$$

where $\mathcal{T}_{k,q}$ is the effective end-to-end channel from transmitter q to receiver k for $k, q \in \mathcal{K}$. Applying a linear receive filter $\mathbf{W}_k \in \mathbb{C}^{N_{\mathrm{R},k} \times d_k}$ to $\mathbf{y}_k$, receiver k obtains

$$\bar{\mathbf{y}}_k = \underbrace{\mathbf{W}_k^{*}\mathcal{T}_{k,k}\mathbf{s}_k}_{\text{desired signal}} + \underbrace{\sum_{\substack{q=1 \\ q \neq k}}^{K} \mathbf{W}_k^{*}\mathcal{T}_{k,q}\mathbf{s}_q}_{\text{interference}} + \underbrace{\sum_{m=1}^{M} \mathbf{W}_k^{*}\mathcal{G}_{k,m}\mathbf{n}_{\mathrm{X},m}}_{\text{enhanced noise from relays}} + \underbrace{\mathbf{W}_k^{*}\mathbf{n}_{\mathrm{R},k}}_{\text{local noise}}. \tag{3.6}$$

The interference-plus-noise covariance matrix at receiver k is

$$\mathbf{R}_k \;=\; \sum_{\substack{q=1\\q\neq k}}^{K} \boldsymbol{\mathcal{T}}_{k,q}\boldsymbol{\mathcal{T}}_{k,q}^{*} + \sum_{m=1}^{M} \sigma_{\mathrm{X},m}^{2}\boldsymbol{\mathcal{G}}_{k,m}\boldsymbol{\mathcal{G}}_{k,m}^{*} + \sigma_{\mathrm{R},k}^{2}\mathbf{I}_{d_k}.$$

For notational convenience in this chapter, I define $\{\mathbf{F}\} \triangleq \{\mathbf{F}_k\}_{k=1}^{K}$, $\{\mathbf{U}\} \triangleq \{\mathbf{U}_m\}_{m=1}^{M}$ and $\{\mathbf{W}\} \triangleq \{\mathbf{W}_k\}_{k=1}^{K}$, which are the main designed variables. I define $\mathbf{F}_{-k} \triangleq \{\mathbf{F}_1, \cdots, \mathbf{F}_{k-1}, \mathbf{F}_{k+1}, \cdots, \mathbf{F}_K\}$ for $k \in \mathcal{K}$. $\{\mathbf{F}\}$ and $(\mathbf{F}_k; \mathbf{F}_{-k})$ can be used interchangeably. I similarly define $\mathbf{U}_{-m}$ for $m \in \mathcal{M}$ and $\mathbf{W}_{-k}$ for $k \in \mathcal{K}$. I define $\boldsymbol{\mathcal{U}}_{m,k} \triangleq \mathbf{U}_m\mathbf{H}_{m,k}$ and $\boldsymbol{\mathcal{W}}_{k,m} \triangleq \mathbf{W}_k^{*}\mathbf{G}_{k,m}$ for $k \in \mathcal{K}$ and $m \in \mathcal{M}$.

Remark 3.2.1. This remarks summarizes the key assumptions in this chapter and their justification.

- *Assumption 3.1: I assume that a stage of multiple-antenna AF relays aid data transmission from the transmitters to the receivers.* The use of multiple antennas allows for the utilization of spatial dimensions at the relays as well as at the transmitters for managing interference.

- *Assumption 3.2: The relay cannot transmit and receive at the same time.* This means I consider only half-duplex relays since they are more practical than full-duplex relays.

- *Assumption 3.3: Each transmitter has data for only one receiver and each receiver is served by only one transmitter.*

- *Assumption 3.4: The transmissions on each hop start at the same time and end at the same time.* This can be done based on the use of a common frame structure. This assumption constraints the types of interference at any time, making analysis more tractable.

- *Assumption 3.5: The channels are frequency-flat, slowly-varying, and block-fading.* For example, the results can be used for a single carrier of MIMO OFDM AF relay interference systems.

- *Assumption 3.6: Direct channels from the transmitters and receivers are ignored by the second-stage receivers.* This assumption helps simplify the analysis. Moreover, it is unclear how to take the advantage of the signals sent directly from the transmitters to the receivers in the relay interference channel. This means that without appropriate signal processing at the receivers, signals received directly from the transmitters may even degrade the system performance.

- *Assumption 3.7: Perfect and instantaneous information of all channels is available at a central processing unit.* Although this is a strict requirement, my results are still valuable since they show the substantial gains that can be achieved through coordination. My results can be used as a benchmark for future work that makes a more practical CSI assumption.

- *Assumption 3.8: Received signals at the relay and receiver are corrupted by additive, circularly symmetric, complex spatially white Gaussian noise. The transmitters, including uncoordinated interferers, use independent*

Gaussian codebooks for transmission. This allows the receivers to treat interference as additive Gaussian noise while decoding their desired signals. This assumption is crucial to deriving the rate expressions used in this chapter. NonGaussian interference and noise are not considered in this chapter.

- *Assumption 3.9: Transmitters use linear transmit precoders and receivers use linear receive filters.* The linear processing at the transmitters and relays are attractive in practice due to its low implementation complexity.

- *Assumption 3.10: The relay interference channel is fully-connected on each hop.* For example, this is reasonable for modeling the scenario where relays are used to provide indoor coverage for a large building.

- *Assumption 3.11: The numbers of data streams and the numbers of antennas satisfy the feasibility conditions of end-to-end interference alignment.* To the best of my knowledge, there have not been any theoretical results on the feasible conditions of end-to-end interference alignment in the MIMO AF relay interference channel.

3.3 Total Leakage Minimization Algorithm

This section presents an interference alignment algorithm for the MIMO AF relay interference channel. It is inspired by the interference alignment algorithms for the single-hop interference channel in [69, 70, 156, 158]. The underlying observation is that when interference alignment is feasible, the sum power of interference at all receivers, also known as sum leakage power, is zero.

3.3.1 Total Interference Plus Enhanced Noise Power Minimization Problem Formulation

From (3.6), there are three groups of undesired signals at each receiver: i) interference, ii) enhanced relay noise, and iii) local noise. Interference depends on $\{\mathbf{F}\}, \{\mathbf{U}\}, \{\mathbf{W}\}$, and the channels on the two hops. Enhanced relay noise depends only on $\{\mathbf{U}\}, \{\mathbf{W}\}$ and the channels on the second hop. Note that enhanced relay noise can be treated as the sum of interference signals at the receivers of a virtual single-hop interference channel with $(M+1)$ users. Let $\mathcal{I}\left(\{\mathbf{F}\}, \{\mathbf{U}\}, \{\mathbf{W}\}\right)$ denote the total interference power of the two-hop AF relay interference channel. By evaluating the expectation, exploiting the independence of transmit signals $\mathbf{s}_k$ for $k \in \mathcal{K}$, and using the equality $\|\mathbf{A}\|_F^2 = \mathrm{tr}(\mathbf{A}\mathbf{A}^*)$, I obtain

$$\begin{aligned} \mathcal{I}\left(\{\mathbf{F}\}, \{\mathbf{U}\}, \{\mathbf{W}\}\right) &= \sum_{k=1}^{K} \sum_{\substack{q=1 \\ q \neq k}}^{K} \mathbb{E} \left\| \mathbf{W}_k^* \boldsymbol{\mathcal{T}}_{k,q} \mathbf{s}_q \right\|_F^2 \\ &= \sum_{k=1}^{K} \sum_{\substack{q=1 \\ q \neq k}}^{K} \mathrm{tr}(\mathbf{W}_k^* \boldsymbol{\mathcal{T}}_{k,q} \boldsymbol{\mathcal{T}}_{k,q}^* \mathbf{W}_k). \end{aligned}$$

Let $\mathcal{N}(\{\mathbf{U}\}, \{\mathbf{W}\})$ denote the sum enhanced relay noise power. By evaluating the expectations and exploiting the independence of the noise vectors at the relays, I obtain

$$\begin{aligned}
\mathcal{N}(\{\mathbf{U}\},\{\mathbf{W}\}) &= \sum_{k=1}^{K}\sum_{m=1}^{M}\mathbb{E}\left\|\mathbf{W}_k^*\mathcal{G}_{k,m}\mathbf{n}_{\mathrm{X},m}\right\|_F^2 \\
&= \sum_{k=1}^{K}\sum_{m=1}^{M}\sigma_{\mathrm{X},m}^2\,\mathrm{tr}(\mathbf{W}_k^*\mathcal{G}_{k,m}\mathcal{G}_{k,m}^*\mathbf{W}_k).
\end{aligned}$$

The total leakage power at all the receivers is defined as $\mathcal{I}(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\}) + \mathcal{N}(\{\mathbf{U}\},\{\mathbf{W}\})$.

It can be argued that the high SNR regime of the relay interference channel corresponds to high transmit power at both the transmitters and relays. Thus, in addition to completely eliminating interference, it is necessary to eliminate enhanced relay noise. Otherwise, enhanced relay noise power scales with the desired signal power, preventing the system from achieving high multiplexing gains. Down scaling the transmit power at either the transmitters or relays further decreases the total leakage power. For example, if $(\{(1/a)\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\})$ is used instead of $(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\})$ where $\{(1/a)\mathbf{F}\} = \{(1/a)\mathbf{F}_1,\cdots,(1/a)\mathbf{F}_K\}$ and $a > 1$, then both the actual transmit power at the transmitters and the total leakage power decrease $a^2 > 1$ times. Therefore, equality power constraints at both the transmitters and relays are required to make it a meaningful design problem. Assuming a sum power constraint at the relays with no power control, I propose to formulate the total leakage minimization problem, denoted as $(\mathcal{TL})$, for the MIMO AF relay interference channel

$$\begin{aligned}
\min_{\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\}} \quad & \mathcal{I}\left(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\}\right)+\mathcal{N}(\{\mathbf{U}\},\{\mathbf{W}\}) \\
\text{subject to} \quad & \operatorname{tr}(\mathbf{F}_k^*\mathbf{F}_k)=p_{\mathrm{T},k}^{\max}, k\in\mathcal{K}, \\
& \sum_{m=1}^{M}\left[\sum_{k=1}^{K}\operatorname{tr}(\mathbf{F}_k^*\mathbf{H}_{m,k}^*\mathbf{U}_m^*\mathbf{U}_m\mathbf{H}_{m,k}\mathbf{F}_k)+\sigma_{\mathrm{X},m}^2\operatorname{tr}(\mathbf{U}_m\mathbf{U}_m^*)\right]=p_{\mathrm{X}}^{\max}.
\end{aligned}$$

Remark 3.3.1. $(\mathcal{TL})$ is nonconvex. In general, finding the globally optimal solutions to $(\mathcal{TL})$ is NP-hard, i.e., it is impossible to find them with reasonable computational complexity.

Remark 3.3.2. $(\mathcal{TL})$ is always feasible. Indeed, $(\mathcal{TL})$ has at least the following feasible solution

$$\begin{aligned}
\mathbf{F}_{0,k} &= \sqrt{\frac{p_{\mathrm{T},k}^{\max}}{d_k}}\mathbf{I}_{N_{\mathrm{T},k}\times d_k}, k\in\mathcal{K}, \\
\mathbf{W}_{0,k} &= \sqrt{\frac{1}{d_k}}\mathbf{I}_{N_{\mathrm{R},k}\times d_k}, k\in\mathcal{K}, \\
\mathbf{U}_{0,m} &= \sqrt{\alpha p_{\mathrm{X}}^{\max}}\mathbf{I}_{N_{\mathrm{X},m}\times N_{\mathrm{X},m}}, m\in\mathcal{M},
\end{aligned}$$

where $\alpha=\left(\sum_{k=1}^{K}\frac{p_{\mathrm{T},k}^{\max}}{d_k}\sum_{m=1}^{M}\operatorname{tr}(\mathbf{H}_{m,k}^*\mathbf{H}_{m,k})+\sum_{m=1}^{M}N_{\mathrm{X},m}\sigma_{\mathrm{X},m}^2\right)^{-1}$. Any feasible solution to $(\mathcal{TL})$ can be used as an initial solution for my algorithm. Note that I claim the feasibility of $(\mathcal{TL})$, but not the feasibility of relay-aided interference alignment.

Remark 3.3.3. The total leakage minimization problem formulated in [145] for an AF relay network is a simplified version of $(\mathcal{TL})$. It is assumed in [145] that the transmitters and receivers are equipped with a single antenna. Each pair is aided by a dedicated multiple-antenna AF relay. The formulation in [145] does not consider power constraints at the relays. It also assumes that there are no crosslinks for the transmissions from relays to receivers, i.e. $\mathbf{G}_{k,q} = \mathbf{0}$ for all $k, q \in \mathcal{K}$ and $k \neq q$. As a result, for fixed $\{\mathbf{F}\}$ and $\{\mathbf{W}\}$, the algorithm in [145] can determine each $\mathbf{U}_m$ independently.

Remark 3.3.4. When perfect end-to-end interference alignment is not possible, the total leakage minimization approach aims at finding as small interference power as possible. This means that this approach does not account for the number of dimensions spanned by interference signals. For the MIMO single-hop interference channel, prior work in [152] shows that the maximization of the multiplexing gains of the MIMO single-hop interference channel is equivalent to a rank-constrained rank minimization problem that minimizes the number of dimensions spanned by interference signals. The idea is to reformulate all interference alignment requirements to requirements involving ranks. Since the rank-constrained rank minimization problem is nonconvex and it is challenging to find its global optima, the authors of [152] propose an alternating minimization algorithm to alternatively design the transmit precoders and the receive filters. Notably, in the transmit precoder design, they have to use convex approximations of the cost function and the rank constraints to heuristically find suboptimal solutions of the rank-constrained rank minimization problem. No theoretical results are provided in [152] to show the tightness of the (heuristic) relaxation. In principle, it is possible to formulate such a rank-constrained rank minimization problem for the MIMO AF relay channel. It is not straightforward to extend the results in [152] to find high quality solutions to the rank-constrained rank minimization problem for the MIMO AF relay channel. For example, the relay processing matrices have complicated effects on desired signal and interference matrices at the receivers, thus it is unclear about appropriate convex approximations for the cost function and constraints in the relay processing matrix design problem.

3.3.2 An Alternating Minimization-based Relay Interference Alignment Algorithm

I adopt an alternating minimization approach to develop an iterative algorithm to find the stationary points of ($\mathcal{TL}$), which is referred to as Algorithm 1. In each iteration, I alternatively fix $(2K + M - 1)$ variables and determine the remaining variable by solving a single-variable optimization problem. The optimization problem in each iteration is always feasible since it has the outcome of the previous iteration as a feasible solution. There are three classes of design subproblems in Algorithm 1: i) receiver filter design, ii) relay processing matrix design, and iii) transmit precoder design. The following subsections present in detail how to solve these subproblems.

3.3.2.1 Receive Filter Design for ($\mathcal{TL}$)

The cost function is rewritten as follows

$$\mathcal{I}(\{\mathbf{F}\}, \{\mathbf{U}\}, \{\mathbf{W}\}) + \mathcal{N}(\{\mathbf{U}\}, \{\mathbf{W}\}) = \sum_{k=1}^{K} \mathrm{tr}(\mathbf{W}_k^* \mathbf{Z}_k \mathbf{W}_k),$$

where $\mathbf{Z}_k = \sum_{\substack{q=1 \\ q \neq k}}^{K} \boldsymbol{\mathcal{T}}_{k,q} \boldsymbol{\mathcal{T}}_{k,q}^* + \sum_{m=1}^{M} \sigma_{\mathrm{X},m}^2 \boldsymbol{\mathcal{G}}_{k,m} \boldsymbol{\mathcal{G}}_{k,m}^*$. Since $\mathbf{W}_k$ for $k \in \mathcal{K}$ are decoupled in (3.7), when $\{\mathbf{F}\}$ and $\{\mathbf{U}\}$ are fixed, I can determine simultaneously each $\mathbf{W}_k$ for $k \in \mathcal{K}$ by solving

$$(\mathcal{TL}\text{-}\mathbf{W}_k) : \mathbf{W}_k = \underset{\mathbf{X} \in \mathbb{C}^{N_{\mathrm{R},k} \times d_k}}{\arg\min} \ \mathrm{tr}(\mathbf{X}^* \mathbf{Z}_k \mathbf{X}).$$

A global optimum of $(\mathcal{TL}\text{-}\mathbf{W}_k)$ is given by [132]

$$\mathbf{W}_k = \nu_{\min}^{d_k}(\mathbf{Z}_k). \tag{3.7}$$

3.3.2.2 Relay Processing Matrix Design for $(\mathcal{TL})$

Fixing $\{\mathbf{F}\}$, $\{\mathbf{W}\}$, and $\mathbf{U}_{-m}$ for a given $m \in \mathcal{M}$, I focus on determining $\mathbf{U}_m$ that minimizes $\mathcal{I}\left(\{\mathbf{F}\}, (\mathbf{U}_{-m}, \mathbf{U}_m), \{\mathbf{W}\}\right) + \mathcal{N}\left((\mathbf{U}_{-m}, \mathbf{U}_m), \{\mathbf{W}\}\right)$. I define

$$\eta_{\mathrm{U},m} \;=\; p_{\mathrm{X}}^{\max} - \sum_{\substack{n=1 \\ n \neq m}}^{M} \sum_{k=1}^{K} \mathrm{tr}\left(\mathbf{U}_n \boldsymbol{\mathcal{H}}_{n,k} \boldsymbol{\mathcal{H}}_{n,k}^* \mathbf{U}_n^*\right) - \sum_{\substack{n=1 \\ n \neq m}}^{M} \sigma_{\mathrm{X},n}^2 \, \mathrm{tr}(\mathbf{U}_n \mathbf{U}_n^*).$$

The single-variable optimization problem for designing $\mathbf{U}_m$ from $(\mathcal{TL})$ is

$$\begin{aligned}
\min_{\mathbf{X} \in \mathbb{C}^{N_{\mathrm{X},m} \times N_{\mathrm{X},m}}} \quad & \sum_{k=1}^{K} \sum_{\substack{q=1 \\ \neq k}}^{K} \mathrm{tr}\left(\mathbf{X} \boldsymbol{\mathcal{H}}_{m,q} \boldsymbol{\mathcal{H}}_{m,q}^* \mathbf{X}^* \boldsymbol{\mathcal{W}}_{k,m}^* \boldsymbol{\mathcal{W}}_{k,m}\right) \\
& + \sigma_{\mathrm{X},m}^2 \sum_{k=1}^{K} \mathrm{tr}(\mathbf{X}^* \boldsymbol{\mathcal{W}}_{k,m}^* \boldsymbol{\mathcal{W}}_{k,m} \mathbf{X}) \\
& + 2\mathbb{R}\left\{ \sum_{k=1}^{K} \sum_{\substack{q=1 \\ q \neq k}}^{K} \sum_{\substack{n=1 \\ n \neq m}}^{M} \mathrm{tr}\left(\mathbf{X} \boldsymbol{\mathcal{H}}_{m,q} \boldsymbol{\mathcal{H}}_{n,q}^* \mathbf{U}_n^* \boldsymbol{\mathcal{W}}_{k,n}^* \boldsymbol{\mathcal{W}}_{k,m}\right) \right\} \\
\text{subject to} \quad & \mathrm{tr}\left(\mathbf{X}\left(\sum_{k=1}^{K} \boldsymbol{\mathcal{H}}_{m,k} \boldsymbol{\mathcal{H}}_{m,k}^* + \sigma_{\mathrm{X},m}^2 \mathbf{I}_{N_{\mathrm{X},m}}\right)\mathbf{X}^*\right) = \eta_{\mathrm{U},m}.
\end{aligned}$$

Because of the special form of the first term in the cost function of $(\mathcal{TL}\text{-}\mathbf{U}_m)$, it is not straightforward to use the methods for the single-hop interference channel like those in [158] to solve $(\mathcal{TL}\text{-}\mathbf{U}_m)$.

I propose a method for transforming ($\mathcal{TL}$-$\mathbf{U}_m$) into an equivalent optimization problem that is more readily solvable. I define a new variable $\mathbf{u}_m = \operatorname{vec}(\mathbf{U}_m) \in \mathbb{C}^{N_{\mathrm{X},m}^2 \times 1}$. $\mathbf{U}_m$ is obtained from $\mathbf{u}_m$ by the vec^{-1} operator. I define the following matrices that are independent of $\mathbf{u}_m$

$$
\begin{aligned}
\mathbf{A}_{1,m} &= \sum_{k=1}^{K} \left(\sum_{\substack{q=1 \\ q \neq k}}^{K} \mathcal{H}_{m,q} \mathcal{H}_{m,q}^{*} + \sigma_{\mathrm{X},m}^{2} \mathbf{I}_{N_{\mathrm{X},m}} \right)^{T} \otimes \left(\mathcal{W}_{k,m}^{*} \mathcal{W}_{k,m} \right), \\
\mathbf{a}_{2,m} &= \operatorname{vec} \left(\sum_{k=1}^{K} \sum_{\substack{q=1 \\ q \neq k}}^{K} \sum_{\substack{n=1 \\ n \neq m}}^{M} \mathcal{W}_{k,m}^{*} \mathcal{W}_{k,n} \mathbf{U}_n \mathcal{H}_{n,q} \mathcal{H}_{m,q}^{*} \right), \\
\mathbf{A}_{3,m} &= \left(\sum_{k=1}^{K} \mathcal{H}_{m,k} \mathcal{H}_{m,k}^{*} + \sigma_{\mathrm{X},m}^{2} \mathbf{I}_{N_{\mathrm{X},m}} \right)^{T} \otimes \mathbf{I}_{N_{\mathrm{X},m}}.
\end{aligned} \tag{3.8}
$$

Note that with probability one, $\mathbf{A}_{3,m}$ is Hermitian and positive definite while $\mathbf{A}_{1,m}$ is Hermitian and positive semidefinite. The equalities $\operatorname{tr}(\mathbf{A}\mathbf{B}\mathbf{A}^*\mathbf{C}) = (\operatorname{vec}(\mathbf{A}))^*(\mathbf{B}^T \otimes \mathbf{C}) \operatorname{vec}(\mathbf{A})$, $\operatorname{tr}(\mathbf{A}^*\mathbf{B}\mathbf{A}) = \operatorname{tr}(\mathbf{A}\mathbf{I}\mathbf{A}^*\mathbf{B}) = (\operatorname{vec}(\mathbf{A}))^*(\mathbf{I} \otimes \mathbf{B}) \operatorname{vec}(\mathbf{A})$, and $\operatorname{tr}(\mathbf{A}\mathbf{B}^*) = (\operatorname{vec}(\mathbf{B}))^* \operatorname{vec}(\mathbf{A})$ [90] are used to transform both the cost function and constraint of ($\mathcal{TL}$-$\mathbf{U}_m$) into quadratic expressions of $\mathbf{u}_m$. Let ($\mathcal{TL}$-$\mathbf{u}_m$) denote the quadratically constrained quadratic program (QCQP) of $\mathbf{u}_m$. It is given by

$$
\begin{aligned}
\min_{\mathbf{x} \in \mathbb{C}^{N_{\mathrm{X},m}^2 \times 1}} \quad & \mathbf{x}^* \mathbf{A}_{1,m} \mathbf{x} + \mathbf{a}_{2,m}^* \mathbf{x} + \mathbf{x}^* \mathbf{a}_{2,m} \\
\text{subject to} \quad & \mathbf{x}^* \mathbf{A}_{3,m} \mathbf{x} = \eta_{\mathrm{U},m}.
\end{aligned} \tag{3.9}
$$

To guarantee convergence, I need to find a global optimum of ($\mathcal{TL}$-$\mathbf{u}_m$). The existence of this global optimum is stated in Proposition 3.3.1.

Proposition 3.3.1. *The problem* ($\mathcal{TL}$-$\mathbf{u}_m$) *has a unique globally optimal solution with probability one.*

Proof. Let θ be the Lagrange multiplier associated with the constraint (3.9). The corresponding Lagrangian function is defined as

$$\mathcal{L}_1(\mathbf{x},\theta) = \mathbf{x}^*(\mathbf{A}_{1,m} + \theta\mathbf{A}_{3,m})\mathbf{x} + \mathbf{a}_{2,m}^*\mathbf{x} + \mathbf{x}^*\mathbf{a}_{2,m} - \theta\eta_{\mathrm{U},m}.$$

Any optimal solution to ($\mathcal{TL}$-$\mathbf{u}_m$) must satisfy the following KKT conditions

$$(\mathbf{A}_{1,m} + \theta\mathbf{A}_{3,m})\mathbf{x} + \mathbf{a}_{2,m} = 0, \tag{3.10}$$

$$\mathbf{x}^*\mathbf{A}_{3,m}\mathbf{x} - \eta_{\mathrm{U},m} = 0. \tag{3.11}$$

Since $\mathbf{A}_{1,m} + \theta\mathbf{A}_{3,m} \neq \mathbf{0}$ with probability one, then it follows from (3.10) that

$$\mathbf{x} = (-1) * (\mathbf{A}_{1,m} + \theta\mathbf{A}_{3,m})^{-1}\mathbf{a}_{2,m}. \tag{3.12}$$

By substituting (3.12) into (3.11), I obtain

$$\mathbf{a}_{2,m}^*(\mathbf{A}_{1,m} + \theta\mathbf{A}_{3,m})^{-1}\mathbf{A}_{3,m}(\mathbf{A}_{1,m} + \theta\mathbf{A}_{3,m})^{-1}\mathbf{a}_{2,m} = \eta_{\mathrm{U},m}. \tag{3.13}$$

I define $\mathcal{Z} \triangleq \{z \in \mathbb{R} : \mathbf{A} + z\mathbf{B} \text{ is a positive definite matrix}\}$. The left-hand side of (3.13) has the form of $g(z) = \mathbf{a}^*(\mathbf{A} + z\mathbf{B})^{-1}\mathbf{B}(\mathbf{A} + z\mathbf{B})^{-1}\mathbf{a}$, where $z \in \mathcal{Z}$, $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{n\times n}$, $\mathbf{A}$ is a positive semidefinitive matrix, $\mathbf{B}$ is a positive semidefinite matrix, and $\mathbf{a} \in \mathbb{C}^{n\times 1}$ is a column vector. Note that there exists a nonsingular matrix $\mathbf{S}$ such that $\mathbf{A} = \mathbf{SCS}^*$ and $\mathbf{B} = \mathbf{SDS}^*$, where $\mathbf{C} = \mathrm{diag}(c_1, \cdots, c_n)$ and $\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_n)$ are the diagonal matrices with $c_i, d_i > 0$ for $i = 1, \cdots, n$ [89]. Note that $g(z)$ is rewritten as follows

$$g(z) = \sum_{i=1}^{n} \frac{d_i}{(c_i + zd_i)^2} y_i^2, \tag{3.14}$$

where $\mathbf{Sa} \triangleq (y_1, \cdots, y_n)$. Thus, $g(z)$ is a monotonically decreasing function of $z \in \mathcal{Z}$, i.e., the left-hand side of (3.13) monotonically decreases in θ. In addition, Remark 3.3.2 ensures that (3.13) always has a solution. Therefore, (3.13) has a unique solution of θ, and equivalently, ($\mathcal{TL}$-$\mathbf{u}_m$) has a unique global optimum. $\square$

Based on the proof of Proposition 3.3.1, I propose a method to compute $\mathbf{U}_m$ in three steps. A simple 1-D search is first performed to find the unique solution θ^* of (3.13). The global optimum of ($\mathcal{TL}$-$\mathbf{u}_m$) is then obtained by substituting θ^* into (3.12). $\mathbf{U}_m$ is finally obtained from $\mathbf{u}_m$ by using $\mathrm{vec}^{-1}()$.

3.3.2.3 Transmit Precoder Design for ($\mathcal{TL}$)

This subsection focuses on designing $\mathbf{F}_k$ for some $k \in \mathcal{K}$ assuming that $\mathbf{F}_{-k}$, $\{\mathbf{U}\}$, and $\{\mathbf{W}\}$ are fixed. I define

$$\eta_{\mathrm{F},k} = p_{\mathrm{X}}^{\max} - \sum_{\substack{q=1 \\ q\neq k}}^{K} \sum_{m=1}^{M} \mathrm{tr}\left(\mathbf{F}_q^* \mathcal{U}_{m,q}^* \mathcal{U}_{m,q} \mathbf{F}_q\right) - \sum_{m=1}^{M} \sigma_{\mathrm{X},m}^2 \,\mathrm{tr}\left(\mathbf{U}_m \mathbf{U}_m^*\right). \quad (3.15)$$

Let ($\mathcal{TL}$-$\mathbf{F}_k$) denote the single-variable optimization problem from ($\mathcal{TL}$) to determine $\mathbf{F}_k$ while conditioning on the other variables. It is written as follows

$$\begin{aligned}
\min_{\mathbf{X}\in\mathbb{C}^{N_{\mathrm{T},k}\times N_{\mathrm{S},k}}} \quad & \mathrm{tr}\left(\mathbf{X}^*\Big(\sum_{\substack{q=1 \\ q\neq k}}^{K}\sum_{m=1}^{M}\sum_{n=1}^{M} \mathcal{U}_{m,k}^* \mathcal{W}_{q,m}^* \mathcal{W}_{q,n} \mathcal{U}_{n,k}\Big)\mathbf{X}\right) \\
\text{subject to} \quad & \mathrm{tr}(\mathbf{X}^*\mathbf{X}) = p_{\mathrm{T},k} \\
& \mathrm{tr}\left(\mathbf{X}^*\Big(\sum_{m=1}^{M} \mathcal{U}_{m,k}^* \mathcal{U}_{m,k}\Big)\mathbf{X}\right) = \eta_{\mathrm{F},k}.
\end{aligned}$$

In general, ($\mathcal{TL}$-$\mathbf{F}_k$) is non-convex and NP-hard. Recall that the counterpart transmit precoder design problem in [158] has a single equality constraint, making it possible to find its globally optimal solution by using the Lagrange multiplier method with simple 1-D search. Nevertheless, ($\mathcal{TL}$-$\mathbf{F}_k$) has two equality constraints, using the Lagrange multiplier method requires a more complicated 2-D search. Thus, ($\mathcal{TL}$-$\mathbf{F}_k$) needs to be solved using another method.

Similarly to Subsection 3.3.2.2, I propose a method for transforming ($\mathcal{TL}$-$\mathbf{F}_k$) into an equivalent optimization problem. I define a new variable $\mathbf{f}_k = \mathrm{vec}(\mathbf{F}_k) \in \mathbb{C}^{N_{\mathrm{T},k} d_k \times 1}$. I also define the following matrices

$$\begin{aligned}
\mathbf{B}_{1,k} &= \mathbf{I}_{d_k} \otimes \left(\sum_{\substack{q=1 \\ q \neq k}}^{K} \sum_{m=1}^{M} \sum_{n=1}^{M} \mathcal{U}_{m,k}^{*} \mathcal{W}_{q,m}^{*} \mathcal{W}_{q,n} \mathcal{U}_{n,k} \right), \\
\mathbf{B}_{2,k} &= \mathbf{I}_{d_k} \otimes \left(\sum_{m=1}^{M} \mathcal{U}_{m,k}^{*} \mathcal{U}_{m,k} \right).
\end{aligned}$$

Both $\mathbf{B}_{1,k}$ and $\mathbf{B}_{2,k}$ are Hermitian positive definite matrices, independent of $\mathbf{f}_k$ and $\mathbf{f}_k^*$. Using $\mathrm{tr}(\mathbf{A}^*\mathbf{B}\mathbf{A}) = (\mathrm{vec}(\mathbf{A}))^*(\mathbf{I} \otimes \mathbf{B})\,\mathrm{vec}(\mathbf{A})$ [90], I transform ($\mathcal{TL}$-$\mathbf{F}_k$) into the following equivalent formulation

$$\begin{aligned}
(\mathcal{TL}\text{-}\mathbf{f}_k) : \mathbf{f}_k = \underset{\mathbf{x} \in \mathbb{C}^{N_{\mathrm{T},k} d_k \times 1}}{\arg\min} \quad & \mathbf{x}^* \mathbf{B}_{1,k} \mathbf{x} \\
\text{subject to} \quad & \mathbf{x}^* \mathbf{x} = p_{\mathrm{T},k}^{\max}, \\
& \mathbf{x}^* \mathbf{B}_{2,k} \mathbf{x} = \eta_{\mathrm{F},k}.
\end{aligned}$$

Note that ($\mathcal{TL}$-$\mathbf{f}_k$) is a complex-valued homogeneous QCQP with two equality quadratic constraints. ($\mathcal{TL}$-$\mathbf{f}_k$) is, however, still non-convex and NP-hard [95, 131].

In solving ($\mathcal{TL}$-$\mathbf{f}_k$), I introduce a new variable $\mathbf{Y} = \mathbf{x}\mathbf{x}^*$. It follows that $\mathbf{Y}$ is a rank-one Hermitian positive semidefinite matrix. In addition, since $\mathbf{a}^*\mathbf{B}\mathbf{a} = \mathrm{tr}(\mathbf{B}\mathbf{a}\mathbf{a}^*)$ for any matrix $\mathbf{B}$ and any vector $\mathbf{a}$ [89], I obtain an equivalent optimization problem of ($\mathcal{TL}$-$\mathbf{f}_k$) as follows

$$
\begin{aligned}
(\mathcal{TL}\text{-}\mathbf{f}_k\mathbf{f}_k^*): \min_{\mathbf{Y}\in\mathbb{C}^{N_{\mathrm{T},k}d_k \times N_{\mathrm{T},k}d_k}} \quad & \mathrm{tr}(\mathbf{B}_{1,k}\mathbf{Y}) \\
\text{subject to} \quad & \mathrm{tr}(\mathbf{Y}) = p_{\mathrm{T},k}^{\max}, \\
& \mathrm{tr}(\mathbf{B}_{2,k}\mathbf{Y}) = \eta_{\mathrm{F},k}, \\
& \mathbf{Y} \succeq \mathbf{0}, \mathrm{rank}(\mathbf{Y}) = 1.
\end{aligned}
$$

While the cost function and all other constraints are convex, the rank constraint is nonconvex. This rank constraint is actually the main difficulty in solving ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$). A relaxed version of ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$) is obtained by dropping this rank constraint. The resulting problem is convex and also known as a semidefinite relaxation (SDR) of ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$). In particular, the relaxation is exact as stated in Proposition 3.3.2. This means that the SDR always has a rank-one global optimum.

Proposition 3.3.2. *The SDR obtained from* ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$) *by relaxing the rank constraint is exact, i.e., it always has a rank-one global optimum.*

Proof. The proof is immediate based on Theorem 3.2. in [95], which states that a complex-valued homogeneous QCQP with n constraints is guaranteed to have a global optimum with rank $r \leq \sqrt{n}$. Therefore, having $n = 2$ constraints, ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$) is guaranteed to have a rank-one global optimum. □

The SDR of ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$) can be solved, to any arbitrary accuracy, in a numerically reliable and efficient manner by readily available (and currently free) software packages, e.g., the convex optimization toolbox CVX [77]. Because the SDR may have general-rank global optima besides its rank-one global optima, it is not guaranteed that solving the SDR by the available software packages provides a desired rank-one global optimum. Fortunately, it is always possible to construct a rank-one global optimum of the SDR from any of its general-rank global optimum, e.g., using the rank reduction procedure in [95], which is an extension of the purification technique in [10]. This procedure also allows the decomposition of the rank-one global optimum, resulting in an expression for $\mathbf{f}_k$ from the resulting rank-one global optimum. Appendix A presents the detailed steps of the rank-reduction procedure using my notation.

3.4 MSE-based Sum-Rate Maximization Algorithm

The algorithm in Section 3.3 uses the space dimensions for minimizing the total leakage power, but it does not take into account the desired signal power. Therefore, it may not perform well in terms of end-to-end sum-rate maximization. In this section, I formulate end-to-end sum-rate maximization problems with and without power control. I then develop algorithms to solve the problems based on a relationship between the end-to-end achievable rates and the MSE values at the receivers. Note that the algorithms find the stationary points of the end-to-end sum-rate maximization problems. The simulation results in Section 3.6 show that the algorithms developed in this section outperform Algorithm 1 at low to medium SNR regimes at the expense of higher algorithmic complexity.

3.4.1 End-to-End Sum-Rate Maximization Problem Formulation

3.4.1.1 Mean Squared Error Computation

From (3.6), the MSE of the estimate of $\mathbf{s}_k$ based on $\bar{\mathbf{y}}_k$ is

$$\begin{aligned} MSE_k &= \mathbb{E}\big(\|\bar{\mathbf{y}}_k - \mathbf{s}_k\|_F^2\big) \quad (3.16)\\ &= \operatorname{tr}\Big(\Big(\mathbf{W}_k^*\boldsymbol{\mathcal{T}}_{k,k} - \mathbf{I}_{d_k}\Big)\Big(\boldsymbol{\mathcal{T}}_{k,k}^*\mathbf{W}_k - \mathbf{I}_{d_k}\Big)\Big) + \sum_{\substack{q=1\\q\neq k}}^{K} \operatorname{tr}\Big(\mathbf{W}_k^*\boldsymbol{\mathcal{T}}_{k,q}\boldsymbol{\mathcal{T}}_{k,q}^*\mathbf{W}_k\Big)\\ &\quad + \sum_{m=1}^{M}\sigma_{\mathrm{X},m}^2 \operatorname{tr}\Big(\mathbf{W}_k^*\boldsymbol{\mathcal{G}}_{k,m}\boldsymbol{\mathcal{G}}_{k,m}^*\mathbf{W}_k\Big) + \sigma_{\mathrm{R},k}^2 \operatorname{tr}\Big(\mathbf{W}_k^*\mathbf{W}_k\Big) \quad (3.17)\\ &= \operatorname{tr}\Big(\mathbf{W}_k^*(\boldsymbol{\mathcal{T}}_{k,k}\boldsymbol{\mathcal{T}}_{k,k}^* + \mathbf{R}_k)\mathbf{W}_k\Big)\\ &\quad - \operatorname{tr}(\mathbf{W}_k^*\boldsymbol{\mathcal{T}}_{k,k}) - \operatorname{tr}(\mathbf{F}_k^*\boldsymbol{\mathcal{T}}_{k,k}^*) + d_k. \quad (3.18) \end{aligned}$$

Note that MSE_k depends on $\{\mathbf{F}\}$, $\{\mathbf{U}\}$, and $\mathbf{W}_k$, but not on $\mathbf{W}_{-k}$. I denote the MSE matrix for receiver k as follows

$$\mathbf{E}_k(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k) = \mathbf{W}_k^*(\boldsymbol{\mathcal{T}}_{k,k}\boldsymbol{\mathcal{T}}_{k,k}^* + \mathbf{R}_k)\mathbf{W}_k - \mathbf{W}_k^*\boldsymbol{\mathcal{T}}_{k,k} - \boldsymbol{\mathcal{T}}_{k,k}^*\mathbf{W}_k + \mathbf{I}_{d_k}.$$

Thus, MSE_k reduces to

$$MSE_k\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k\big) = \operatorname{tr}\Big(\mathbf{E}_k\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k\big)\Big). \quad (3.19)$$

It follows from (3.16) and (3.19) that $\mathbf{E}_k$ is a Hermitian and positive semidefinite matrix for $k \in \mathcal{K}$.

The receive filter $\mathbf{W}_k$ that minimizes $MSE_k\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k\big)$ can be determined based on the gradient of MSE_k with respect to $\mathbf{W}_k^*$, which is given by

$$\frac{\partial MSE_k}{\partial \mathbf{W}_k^*} = \mathbf{W}_k(\boldsymbol{\mathcal{T}}_{k,k}\boldsymbol{\mathcal{T}}_{k,k}^* + \mathbf{R}_k) - \boldsymbol{\mathcal{T}}_{k,k}. \tag{3.20}$$

By solving $\frac{\partial MSE_k}{\partial \mathbf{W}_k^*} = 0$, I obtain the linear MMSE receive filter

$$\mathbf{W}_k^{\mathrm{MMSE}} = (\boldsymbol{\mathcal{T}}_{k,k}\boldsymbol{\mathcal{T}}_{k,k}^* + \mathbf{R}_k)^{-1}\boldsymbol{\mathcal{T}}_{k,k}. \tag{3.21}$$

Let $\mathbf{E}_k^{\mathrm{MMSE}}$ be the MSE matrix corresponding to the use of $\mathbf{W}_k^{\mathrm{MMSE}}$, which is a function of $(\{\mathbf{F}\},\{\mathbf{U}\})$. By substituting (3.21) into (3.19), and applying the binomial inverse theorem [89], it follows that

$$\begin{aligned}\mathbf{E}_k^{\mathrm{MMSE}} &= \mathbf{I}_{d_k} - \boldsymbol{\mathcal{T}}_{k,k}^*\big(\boldsymbol{\mathcal{T}}_{k,k}\boldsymbol{\mathcal{T}}_{k,k}^* + \mathbf{R}_k\big)^{-1}\boldsymbol{\mathcal{T}}_{k,k} \\ &= \big(\mathbf{I}_{d_k} + \boldsymbol{\mathcal{T}}_{k,k}^*\mathbf{R}_k^{-1}\boldsymbol{\mathcal{T}}_{k,k}\big)^{-1}. \end{aligned} \tag{3.22}$$

3.4.1.2 Sum-Rate Maximization Problem Formulation

For tractable analysis, I assume Gaussian signaling is used in the system. Thus the maximum achievable rate for the transmission from transmitter k to receiver k when a linear receive filter is used at receiver k is given by

$$\begin{aligned}R_k\big(\{\mathbf{F}\},\{\mathbf{U}\}\big) &= \log_2\det\big(\mathbf{I}_{d_k} + \boldsymbol{\mathcal{T}}_{k,k}^*\mathbf{R}_k^{-1}\boldsymbol{\mathcal{T}}_{k,k}\big) \\ &= -\log_2\det\Big(\mathbf{E}_k^{\mathrm{MMSE}}\big(\{\mathbf{F}\},\{\mathbf{U}\}\big)\Big).\end{aligned}$$

The sum of end-to-end achievable rates is defined as

$$R_{\mathrm{sum}}\big(\{\mathbf{F}\},\{\mathbf{U}\}\big) = -\sum_{k=1}^{K}\log_2\det\Big(\mathbf{E}_k^{\mathrm{MMSE}}(\{\mathbf{F}\},\{\mathbf{U}\})\Big).$$

For consistency with ($\mathcal{TL}$), I consider first the end-to-end sum-rate maximization problem without power control, which is denoted asd ($\mathcal{SR}$-EQ) and formulated as follows

$$\begin{aligned}
\min_{\{\mathbf{F}\},\{\mathbf{U}\}} \quad & -R_{\text{sum}}\big(\{\mathbf{F}\},\{\mathbf{U}\}\big) \\
\text{subject to} \quad & \operatorname{tr}\left(\mathbf{F}_k\mathbf{F}_k^*\right) = p_{\mathrm{T},k}^{\max}, k = 1,\cdots,K \\
& \sum_{m=1}^{M}\sum_{k=1}^{K}\operatorname{tr}\left(\mathbf{U}_m\mathbf{H}_{m,k}\mathbf{F}_k\mathbf{F}_k^*\mathbf{H}_{m,k}^*\mathbf{U}_m^*\right) \\
& + \sum_{m=1}^{M}\sigma_{\mathrm{X},m}^2\operatorname{tr}\left(\mathbf{U}_m\mathbf{U}_m^*\right) = p_{\mathrm{X}}^{\max}.
\end{aligned}$$

Note that ($\mathcal{SR}$-EQ) is nonconvex and NP-hard [124], i.e., its global optima cannot be found efficiently in terms of computational complexity. Thus, in this chapter, I develop an algorithm for finding the stationary points of ($\mathcal{SR}$-EQ) with reasonable computational complexity.

Remark 3.4.1. Replacing the equality constraints of ($\mathcal{SR}$-EQ) by inequality constraints results in the end-to-end sum-rate maximization problem with power control. The problem with inequality constraints is referred to as ($\mathcal{SR}$-NEQ).

Remark 3.4.2. It is challenging, even infeasible with widely available computing systems, to find the global optima of either ($\mathcal{SR}$-EQ) or ($\mathcal{SR}$-NEQ) using exhaustive search. Consider a small MIMO AF relay system with two transmitters, two relays, and two receivers. Each node has two antennas. Given that the receivers always use the linear MMSE receive filters, the sum-rate maximization problem for this small system requires the determination of twelve complex numbers for the transmit precoders and relay processing matrices. Exhaustive search for this combinatorial optimization problem is challenging with high precision. Moreover, in Section 3.6, I provide simulation results that give some hints on the gap between the solutions found by my proposed algorithms and the optimal approach. Based on the results, it is suspected that the gap is not too large.

3.4.1.3 Linking Sum-Rate Maximization and Weighted Sum-MSE Minimization Problems

Note that R_{sum} can be expressed as a function of the MSE matrices at all the receivers. Similar expressions of sum-rates as functions of MSE matrices are obtained for other systems such as the MIMO broadcast channel [48], MIMO interference broadcast channel [165], and two-way relay channel [119, 221]. This relationship makes it possible to formulate a matrix-weighted sum-MSE minimization problem that has the same optimal solutions as $(\mathcal{SR})$. I introduce auxiliary weight matrix variables $\{\mathbf{V}\} \triangleq (\mathbf{V}_1, \cdots, \mathbf{V}_K)$ with $\mathbf{V}_k \in \mathbb{C}^{d_k \times d_k}$ for $k \in \mathcal{K}$. The matrix-weighted sum-MSEs is defined as follows

$$\begin{aligned} &WMSE_{\text{sum}}(\{\mathbf{F}\}, \{\mathbf{U}\}, \{\mathbf{W}\}, \{\mathbf{V}\}) = \\ &\qquad \sum_{k=1}^{K} \Bigg(\operatorname{tr}\Big(\mathbf{V}_k \mathbf{E}_k\big(\{\mathbf{F}\}, \{\mathbf{U}\}, \{\mathbf{W}\}\big)\Big) - \log_2 \det\big(\mathbf{V}_k\big)\Bigg). \end{aligned}$$

I formulate the following matrix-weighted sum-MSE minimization problem, denoted as ($\mathcal{WMSE}$-EQ)

$$
\begin{aligned}
\min_{\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\},\{\mathbf{V}\}} \quad & WMSE_{\text{sum}}\big(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\},\{\mathbf{V}\}\big) \\
\text{subject to} \quad & \operatorname{tr}(\mathbf{F}_k\mathbf{F}_k^*) = p_{\mathrm{T},k}^{\max}, k \in \mathcal{K} \\
& \sum_{m=1}^{M}\sum_{k=1}^{K} \operatorname{tr}(\mathbf{U}_m\mathbf{H}_{m,k}\mathbf{F}_k\mathbf{F}_k^*\mathbf{H}_{m,k}^*\mathbf{U}_m^*) \\
& + \sum_{m=1}^{M} \sigma_{\mathrm{X},m}^2 \operatorname{tr}(\mathbf{U}_m\mathbf{U}_m^*) = p_{\mathrm{X}}^{\max}.
\end{aligned}
$$

Note that $\{\mathbf{V}\}$ appears only in the objective function of ($\mathcal{WMSE}$-EQ). The key observation for solving ($\mathcal{SR}$-EQ) via solving ($\mathcal{WMSE}$-EQ) is stated in Proposition 3.4.1.

Proposition 3.4.1. *($\mathcal{WMSE}$-EQ) is equivalent to ($\mathcal{SR}$-EQ) in the sense that they have the same stationary points.*

Proof. In ($\mathcal{WMSE}$-EQ), $\mathbf{W}_k$ appears only in $\operatorname{tr}(\mathbf{V}_k\mathbf{E}_k(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\}))$ in the objective function. When the other parameters are fixed, it follows that the optimal linear receive filter for ($\mathcal{WMSE}$-EQ) is also $\mathbf{W}_k^{\text{MMSE}}$, which is given in (3.21). Since ($\mathcal{WMSE}$-EQ) and ($\mathcal{SR}$-EQ) have the same constraints, then to check their equivalence, it is sufficient to check if the differentials of their objective functions are the same. These differentials are given by

$$
\begin{aligned}
\mathrm{d}(-R_{\text{sum}}) &= \sum_{k=1}^{K} \operatorname{tr}(\mathbf{E}_k^{-1}\mathrm{d}\mathbf{E}_k), \\
\mathrm{d}WMSE_{\text{sum}} &= \sum_{k=1}^{K} \operatorname{tr}(\mathbf{V}_k\mathrm{d}\mathbf{E}_k) + \sum_{k=1}^{K}\Big(\operatorname{tr}(\mathbf{E}_k\mathrm{d}\mathbf{V}_k) - \operatorname{tr}(\mathbf{V}_k^{-1}\mathrm{d}\mathbf{V}_k)\Big).
\end{aligned}
$$

For any fixed $\{\mathbf{F}\}$, $\{\mathbf{U}\}$, and $\{\mathbf{W}\}$, then $\mathbf{E}_k\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k\big)$ is independent of $\{\mathbf{V}\}$ for $k \in \mathcal{K}$. The optimal matrix weights $\{\mathbf{V}^{\text{opt}}\}_{k=1}^{K}$ of ($\mathcal{WMSE}$-EQ) are given by

$$\begin{aligned} \mathbf{V}_k^{\text{opt}}\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k^{\text{MMSE}}\big) &= \mathbf{E}_k^{-1}\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k^{\text{MMSE}}\big), \\ &= \mathbf{I}_{d_k} + \mathfrak{T}_{k,k}^{*}\mathbf{R}_k^{-1}\mathfrak{T}_{k,k}. \end{aligned} \tag{3.23}$$

Thus, if $\mathbf{V}_k = \mathbf{V}_k^{\text{opt}}\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k^{\text{MMSE}}\big)$ for $k \in \mathcal{K}$, then I have $\text{tr}(\mathbf{E}_k \text{d}\mathbf{V}_k) - \text{tr}(\mathbf{V}_k^{-1}\text{d}\mathbf{V}_k) = 0$ for $k \in \mathcal{K}$, leading to $\text{d}(-R_{\text{sum}}) = \text{d}WMSE_{\text{sum}}$. □

Proposition 3.4.1 states that any stationary point of ($\mathcal{WMSE}$-EQ) is also a stationary point of ($\mathcal{SR}$-EQ) and vice versa. Therefore, it is possible to find the stationary points of ($\mathcal{SR}$) indirectly via solving ($\mathcal{WMSE}$-EQ) rather than directly solving $\mathcal{SR}$.

Remark 3.4.3. The weighted sum-MSE value $WMSE_{\text{sum}}(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\},\{\mathbf{V}\})$ is convex with respect to $\mathbf{F}_k$ for $k \in \mathcal{K}$ if the matrix weights are always chosen as $\mathbf{V}_k = \mathbf{V}_k^{\text{opt}}\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k^{\text{MMSE}}\big)$ according to (3.23). Indeed, it follows from (3.17) that MSE_k is convex with respect to $\mathbf{F}_q$ for all $k, q \in \mathcal{K}$. By construction, $\mathbf{V}_k = \mathbf{V}_k^{\text{opt}}\big(\{\mathbf{F}\},\{\mathbf{U}\},\mathbf{W}_k^{\text{MMSE}}\big)$ is a Hermitian and positive semidefinite matrix for $k \in \mathcal{K}$. By definition, $WMSE_{\text{sum}}(\{\mathbf{F}\},\{\mathbf{U}\},\{\mathbf{W}\},\{\mathbf{V}\})$ is also convex with respect to $\mathbf{F}_k$ for $k \in \mathcal{K}$.

3.4.2 An MSE-based Algorithm for End-to-End Sum-Rate Maximization without Power Control

In this subsection, I propose an algorithm for solving ($\mathcal{WMSE}$-EQ). Specifically, adopting an alternating minimization approach, I develop an algorithm for finding the stationary points of ($\mathcal{WMSE}$-EQ) as well as ($\mathcal{SR}$-EQ), which is referred to as Algorithm 2. The design subproblems in the iterations of Algorithm 2 belong to one of the following four categories.

3.4.2.1 Matrix Weight Design for ($\mathcal{SR}$-EQ)

Since the matrix weights $\mathbf{V}_k^{\text{opt}}$ for $k \in \mathcal{K}$ are independent of each other, they can be updated simultaneously based on (3.19) and (3.23) for a given $\{\mathbf{F}\}$, $\{\mathbf{U}\}$, and $\{\mathbf{W}\}$.

3.4.2.2 Receive Filter Design for ($\mathcal{SR}$-EQ)

As discussed in the proof of Proposition 3.4.1, the optimal solution of ($\mathcal{WMSE}$-EQ) requires that the receivers use the linear MMSE receive filters $\mathbf{W}_k^{\text{MMSE}}$ given in (3.21). Similar to Algorithm 1, the receive filters can be updated simultaneously for a given $\{\mathbf{F}\}$, $\{\mathbf{U}\}$, and $\{\mathbf{V}\}$.

3.4.2.3 Relay Processing Matrix Design for ($\mathcal{SR}$-EQ)

I need to determine $\mathbf{U}_m$ for some $m \in \mathcal{M}$ assuming that $\{\mathbf{F}\}$, $\{\mathbf{W}\}$, $\{\mathbf{V}\}$, and $\mathbf{U}_{-m}$ are fixed. The single-variable optimization problem for designing $\mathbf{U}_m$ from ($\mathcal{WMSE}$-EQ) is denoted as ($\mathcal{WMSE}$-EQ-$\mathbf{U}_m$). It is obtained by performing algebraic manipulations and using $\eta_{\text{U},m}$ as defined in (3.8). It can be expressed as follows

$$\begin{aligned}
\min_{\mathbf{X}\in\mathbb{C}^{N_{\mathrm{X},m}\times N_{\mathrm{X},m}}} \quad & \sum_{k=1}^{K}\sum_{\substack{q=1\\q\neq k}}^{K}\mathrm{tr}\left(\mathbf{X}\boldsymbol{\mathcal{H}}_{m,q}\boldsymbol{\mathcal{H}}_{m,q}^{*}\mathbf{X}^{*}\boldsymbol{\mathcal{W}}_{k,m}^{*}\mathbf{V}_{k}\boldsymbol{\mathcal{W}}_{k,m}\right) \\
& + \sigma_{\mathrm{X},m}^{2}\sum_{k=1}^{K}\mathrm{tr}(\mathbf{X}^{*}\boldsymbol{\mathcal{W}}_{k,m}^{*}\mathbf{V}_{k}\boldsymbol{\mathcal{W}}_{k,m}\mathbf{X}) \\
& -2\mathbb{R}\left\{\sum_{k=1}^{K}\sum_{\substack{q=1\\q\neq k}}^{K}\sum_{\substack{n=1\\n\neq m}}^{M}\mathrm{tr}(\mathbf{X}\boldsymbol{\mathcal{H}}_{m,q}\boldsymbol{\mathcal{H}}_{n,q}^{*}\mathbf{U}_{n}^{*}\boldsymbol{\mathcal{W}}_{k,n}^{*}\mathbf{V}_{k}\boldsymbol{\mathcal{W}}_{k,m})\right\} \\
\text{subject to} \quad & \mathrm{tr}\left(\mathbf{X}^{*}\Big(\sum_{k=1}^{K}\boldsymbol{\mathcal{H}}_{m,k}\boldsymbol{\mathcal{H}}_{m,k}^{*}+\sigma_{\mathrm{X},m}^{2}\mathbf{I}_{N_{\mathrm{X},m}}\Big)\mathbf{X}\right)=\eta_{\mathrm{U},m}.
\end{aligned}$$

Note that ($\mathcal{WMSE}$-EQ-$\mathbf{U}_m$) differs from ($\mathcal{TL}$-$\mathbf{U}_m$) mainly due to the appearance of $\mathbf{V}_k$ in the cost function. Thus, the same steps in Subsection 3.3.2.2 are used to develop the method for finding $\mathbf{U}_m$.

In solving ($\mathcal{WMSE}$-EQ-$\mathbf{U}_m$), I define a new variable $\mathbf{u}_m = \mathrm{vec}(\mathbf{U}_m)$. I also define the following matrices that are independent of $\mathbf{u}_m$ and $\mathbf{u}_m^*$

$$\begin{aligned}
\mathbf{C}_{1,m} &= \sum_{k=1}^{K}\left(\sum_{q=1}^{K}\boldsymbol{\mathcal{H}}_{m,q}\boldsymbol{\mathcal{H}}_{m,q}^{*}+\sigma_{\mathrm{X},m}^{2}\mathbf{I}_{N_{\mathrm{X},m}}\right)^{T}\otimes\left(\boldsymbol{\mathcal{W}}_{k,m}^{*}\mathbf{V}_{k}\boldsymbol{\mathcal{W}}_{k,m}\right), \\
\mathbf{c}_{2,m} &= \mathrm{vec}\left(-\sum_{k=1}^{K}\boldsymbol{\mathcal{W}}_{k,m}^{*}\mathbf{V}_{k}\boldsymbol{\mathcal{H}}_{m,k}^{*}+\sum_{\substack{n=1\\n\neq m}}^{M}\sum_{k=1}^{K}\sum_{q=1}^{K}\boldsymbol{\mathcal{W}}_{k,m}^{*}\mathbf{V}_{k}\boldsymbol{\mathcal{W}}_{k,n}\mathbf{U}_{n}\boldsymbol{\mathcal{H}}_{n,q}\boldsymbol{\mathcal{H}}_{m,q}^{*}\right).
\end{aligned}$$

Using the same manipulations as in Subsection 3.3.2.2 and denoting $\mathbf{C}_{3,m} = \mathbf{A}_{3,m}$, defined in (3.8), I obtain an equivalent formulation of ($\mathcal{WMSE}$-EQ-$\mathbf{U}_m$) for determining $\mathbf{u}_m$, which is referred to as ($\mathcal{WMSE}$-EQ-$\mathbf{u}_m$). It is expressed as follows

$$\begin{aligned} \mathbf{u}_m = \underset{\mathbf{x}\in\mathbb{C}^{N_{\mathrm{X},m}^2\times 1}}{\arg\min} \quad & \mathbf{x}^*\mathbf{C}_{1,m}\mathbf{x} + \mathbf{c}_{2,m}^*\mathbf{x} + \mathbf{x}^*\mathbf{c}_{2,m} \\ \text{subject to} \quad & \mathbf{x}^*\mathbf{C}_{3,m}\mathbf{x} = \eta_{\mathrm{U},m}. \end{aligned} \tag{3.24}$$

Note that ($\mathcal{WMSE}$-EQ-$\mathbf{u}_m$) has the same form as ($\mathcal{TL}$-$\mathbf{u}_m$), thus its global optimum can be found using the Lagrange multiplier method. I first find the positive Lagrange multiplier β^* associated with the constraint (3.24) by solving the following equation

$$\mathbf{c}_{2,m}^*(\mathbf{C}_{1,m} + \beta\mathbf{C}_{3,m})^{-1}\mathbf{C}_{3,m}(\mathbf{C}_{1,m} + \beta\mathbf{C}_{3,m})^{-1}\mathbf{c}_{2,m} = \eta_{\mathrm{U},m}.$$

I then obtain the global optimum of ($\mathcal{WMSE}$-EQ-$\mathbf{u}_m$) as $\mathbf{u}_m = -(\mathbf{C}_{1,m} + \beta\mathbf{C}_{3,m})^{-1}\mathbf{c}_{2,m}$. I finally use the vec^{-1} operator to get $\mathbf{U}_m$ from $\mathbf{u}_m$.

3.4.2.4 Transmit Precoder Design for ($\mathcal{SR}$-EQ)

Assuming $\{\mathbf{W}\}$, $\{\mathbf{U}\}$, $\{\mathbf{V}\}$, and $\mathbf{F}_{-k}$ are fixed for some $k \in \mathcal{K}$, I focus on determining $\mathbf{F}_k$. For notational convenience, I define the following matrices

$$
\begin{aligned}
\mathbf{D}_{1,k} &= \sum_{q=1}^{K}\sum_{m=1}^{M}\sum_{n=1}^{M} \boldsymbol{\mathcal{U}}_{m,k}^{*}\boldsymbol{\mathcal{W}}_{q,m}^{*}\mathbf{V}_q\boldsymbol{\mathcal{W}}_{q,n}\boldsymbol{\mathcal{U}}_{n,k},\\
\mathbf{D}_{2,k} &= \sum_{m=1}^{M} \boldsymbol{\mathcal{U}}_{m,k}^{*}\boldsymbol{\mathcal{W}}_{k,m}^{*}\mathbf{V}_k^{*},\\
\mathbf{D}_{3,k} &= \sum_{m=1}^{M} \boldsymbol{\mathcal{U}}_{m,k}^{*}\boldsymbol{\mathcal{U}}_{m,k},\\
\mathbf{D}_{4,k} &= \sum_{q=1}^{K}\mathbf{V}_q + \sum_{\substack{q=1\\q\neq k}}^{K}\sum_{\substack{p=1\\p\neq k}}^{K}\mathbf{V}_q\mathbf{W}_q^{*}\boldsymbol{\mathcal{T}}_{q,p}\boldsymbol{\mathcal{T}}_{q,p}^{*}\mathbf{W}_q - 2\mathbb{R}\Bigg(\sum_{\substack{q=1\\q\neq k}}^{K}\mathbf{V}_q\mathbf{W}_q^{*}\boldsymbol{\mathcal{T}}_{q,q}\Bigg)\\
&\quad + \sum_{q=1}^{K}\sum_{m=1}^{M}\sigma_{\mathrm{X},m}^{2}\mathbf{V}_q\mathbf{W}_q^{*}\boldsymbol{\mathcal{G}}_{q,m}\boldsymbol{\mathcal{G}}_{q,m}^{*}\mathbf{W}_q + \sum_{q=1}^{K}\sigma_{\mathrm{R},q}^{2}\mathbf{V}_q\mathbf{W}_q\mathbf{W}_q^{*}.
\end{aligned}
$$

After some manipulation of ($\mathcal{WMSE}$-EQ) and with $\eta_{\mathrm{T},k}$ defined in (3.15), I need to solve the following optimization problem ($\mathcal{WMSE}$-EQ-$\mathbf{F}_k$) for $\mathbf{F}_k$

$$
\begin{aligned}
\mathbf{F}_k = \underset{\mathbf{X}\in\mathbb{C}^{N_{\mathrm{T},k}\times d_k}}{\arg\min} \quad & \mathrm{tr}(\mathbf{X}^{*}\mathbf{D}_{1,k}\mathbf{X}) - \mathrm{tr}(\mathbf{D}_{2,k}^{*}\mathbf{X}) - \mathrm{tr}(\mathbf{D}_{2,k}\mathbf{X}^{*}) + \mathrm{tr}(\mathbf{D}_{4,k})\\
\text{subject to} \quad & \mathrm{tr}(\mathbf{X}^{*}\mathbf{X}) = p_{\mathrm{T}}^{\max},\\
& \mathrm{tr}(\mathbf{X}^{*}\mathbf{D}_{3,k}\mathbf{X}) = \eta_{\mathrm{T},k}.
\end{aligned}
$$

I introduce a new variable $\mathbf{Y} = \begin{pmatrix}\mathbf{f}_k\\ 1\end{pmatrix}\begin{pmatrix}\mathbf{f}_k^{*} & 1\end{pmatrix} = \begin{pmatrix}\mathbf{f}_k\mathbf{f}_k^{*} & \mathbf{f}_k\\ \mathbf{f}_k^{*} & 1\end{pmatrix}$, where $\mathbf{f}_k = \mathrm{vec}(\mathbf{F}_k)$. It follows that $\mathbf{Y}$ is a rank-one Hermitian positive semidefinite matrix with the bottom right entry equal to one. Next, ($\mathcal{WMSE}$-EQ-$\mathbf{F}_k$) is transformed into an equivalent formulation, which is denoted as ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^{*}$).

It is expressed as follows

$$
\begin{aligned}
\min_{\mathbf{Y}\in\mathbb{C}^{(N_{\mathrm{T},k}d_k+1)\times(N_{\mathrm{T},k}d_k+1)}} \quad & \mathrm{tr}\left(\begin{pmatrix} \mathbf{I}_{d_k}\otimes\mathbf{D}_{1,k} & -\operatorname{vec}(\mathbf{D}_{2,k}) \\ -(\operatorname{vec}(\mathbf{D}_{2,k}))^* & 1 \end{pmatrix}\mathbf{Y}\right) \\
\text{subject to} \quad & \mathrm{tr}(\mathbf{Y}) = p_{\mathrm{T}}^{\max}+1, \\
& \mathrm{tr}\left(\begin{pmatrix} \mathbf{I}_{d_k}\otimes\mathbf{D}_{3,k} & \mathbf{0}_{N_{\mathrm{T},k}d_k\times 1} \\ \mathbf{0}_{1\times N_{\mathrm{T},k}d_k} & 1 \end{pmatrix}\mathbf{Y}\right) = \eta_{\mathrm{T},k}+1, \\
& \begin{pmatrix} \mathbf{0}_{N_{\mathrm{T},k}d_k\times N_{\mathrm{T},k}d_k} & \mathbf{0}_{N_{\mathrm{T},k}d_k\times 1} \\ \mathbf{0}_{1\times N_{\mathrm{T},k}d_k} & 1 \end{pmatrix}\mathbf{Y} = 1, \\
& \mathbf{Y}\succeq\mathbf{0}, \mathrm{rank}(\mathbf{Y}) = 1.
\end{aligned}
$$

Similar to the methodology for solving ($\mathcal{TL}$-$\mathbf{f}_k\mathbf{f}_k^*$), I adopt the SDP method for finding a global optimum of ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^*$). Specifically, an SDR of ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^*$) is obtained by dropping the non convex rank constraint. This SDR is a convex optimization problem. More importantly, since ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^*$) has $n = 3$ constraints (excluding the rank-one constraint), the relaxation is exact. This means that the SDR of ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^*$) always has a rank-one global optimum. Thus, it is possible to find a general-rank global optimum of the SDR of ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^*$) using readily available software packages, e.g., the CVX toolbox. A rank-one global optimum of the SDR is then constructed from the resulting general-rank global optimum using the rank-reduction procedure in [95]. Note that the last entry of the column vector obtained by the decomposition of the rank-one global optimum may be a complex number with modulus of one. By multiplying the resulting column vector with the conjugate of its last entry, I obtain a desired column vector in the form of $(\mathbf{f}_k \ 1)^T$, which corresponds to another rank-one global optimum of ($\mathcal{WMSE}$-EQ-$\mathbf{f}_k\mathbf{f}_k^*$). Appendix A presents the detailed steps of the rank-reduction procedure using my notation.

3.4.3 An MSE-based Algorithm for End-to-End Sum-Rate Maximization with Power Control

In this subsection, I develop an algorithm for solving ($\mathcal{SR}$-NEQ), which is also based on the relationship between achievable end-to-end rates and MSE values. Similar to Subsection 3.4.1.3, I formulate the corresponding matrix-weighted sum-MSE minimization problem that has the same stationary points as ($\mathcal{SR}$-NEQ). It is referred to as ($\mathcal{WMSE}$-NEQ). The same steps as those in Subsection 3.4.2 are used to develop an alternating minimization algorithm for finding the stationary points of ($\mathcal{SR}$-NEQ). This algorithm is referred to as Algorithm 3. To reduce redundancies, I only compare and contrast the steps of Algorithm 3 and those of Algorithm 2 in this section. The details of Algorithm 3 are provided in [207]. First, the matrix weight and receive filter designs for ($\mathcal{SR}$-NEQ) are exactly the same as those for ($\mathcal{SR}$-EQ). Second, the relay processing matrix design for ($\mathcal{SR}$-NEQ) can be solved using the Lagrangian multiplier method with the only difference being that the multiplier must be nonnegative. Finally, the optimization problem for the transmit precoder design for ($\mathcal{SR}$-NEQ) is obtained by replacing the equality constraints in ($\mathcal{WMSE}$-EQ-$\mathbf{F}_k$) by the corresponding inequality constraints. Fortunately, the resulting optimization problem is convex with respect to $\mathbf{F}_k$. In particular, it follows from Remark 3.4.3 that the objective function of the problem ($\mathcal{WMSE}$-$\mathcal{F}_k$) is convex with respect to $\mathbf{F}_k$. In addition, since $\mathbf{D}_{3,k}$ is a Hermitian and positive semidefinite matrix, then it follows that the constraints of the resulting problem are also convex with respect to $\mathbf{F}_k$. Thus, any available software package for convex optimization like the CVX toolbox [77] could be used to solve for its unique global optimum.

3.5 Discussion

In this section, I discuss the proposed algorithms from the following aspects: i) their convergence, ii) the quality of the solution, iii) the assumption on power constraints at the relays, and iv) the impacts of transmitter topology.

In terms of convergence, all the proposed algorithms are guaranteed to converge to a stationary point of the corresponding multi-variable optimization problem. Furthermore, they are based on alternating minimization. In each iteration, the objective is tominimize the cost function of the original multi-variable optimization problem, which is either the total leakage power in ($\mathcal{TL}$) or the matrix-weighted sum-MSEs in ($\mathcal{WMSE}$-EQ) and ($\mathcal{WMSE}$-NEQ). Since a global optimum of the corresponding single-variable minimization problem can be found at each iteration, the cost function of the original multi-variable optimization problem is non-increasing after each iteration.

The proposed algorithms are not guaranteed to reach a global optimum of the corresponding multi-variable optimization problem. Thus, the quality of their solution depends significantly on the initial solution selected in the first iteration. One way to improve the proposed algorithms is to run them multiple times with different initializations and select the solution that provides the best objective value. Nevertheless, this comes at the expense of longer running time.

The assumption on power constraints at the relays is mainly because the method is applied to solve the transmit precoder design problems in Algorithm 1 and Algorithm 2. Specifically, I propose to use the SDP method to find a global optimum of the transmit precoder design problems, which are nonconvex complex-valued homogeneous QCQPs with equality constraints. Prior work in [10, 95] shows that the relaxed SDR of the QCQPs is guaranteed to have a rank-one global optimum if there are at most four equality quadratic constraints. The sum power constraint at the relays in either ($\mathcal{TL}$) or ($\mathcal{SR}$-EQ), and hence ($\mathcal{WMSE}$-EQ), is crucial since it results in only two quadratic equality constraints. Note that with per-relay power constraints, the number of quadratic constraints of the resulting QCQP is $(M+1)$. Therefore, when power control is not considered, the similar total leakage minimization problem and end-to-end sum-rate maximization problem with per-relay power constraints can be formulated using the same steps if there are at most three relays, i.e., $M \leq 3$. The sum power constraint at the relays, however, is not crucial to the formulation of ($\mathcal{SR}$-NEQ), and hence that of ($\mathcal{WMSE}$-NEQ), and Algorithm 3. For the end-to-end sum-rate maximization problem with power control, regardless of the type of power constraints at the relays, the single-variable optimization problems for designing the relay processing matrices and the transmit precoders are convex with respect to the corresponding variable [207]. Thus, the global optimum can be found at each iteration.

In cellular networks, the algorithms work for any fixed topology of base stations, not just the hexagonal grid model of base station locations. As discussed in Remark 2.6.1, the actual deployment of cellular networks leads to more random base station locations [220] and a more tractable model for base station topology is proposed [13]. Future work may investigate the impact of interference, including uncoordinated interference, on my proposed algorithms by using this framework.

3.6 Simulations

This section presents Monte Carlo simulation results to investigate the average end-to-end sum-rate performance and to gain insights into the multiplexing gains achieved by the proposed algorithms. I consider only symmetric relay systems, which are denoted as $(N_\mathrm{R} \times N_\mathrm{T}, d)^K + N_\mathrm{X}^M$, where $N_{\mathrm{R},k} = N_\mathrm{R}$, $N_{\mathrm{T},k} = N_\mathrm{T}$, $d_k = d$ and $N_{\mathrm{X},m} = N_\mathrm{X}$ for $k \in \mathcal{K}, m \in \mathcal{M}$. The power values are normalized such that $\sigma_{\mathrm{R},k} = \sigma_{\mathrm{X},m} = 1$, $p_{\mathrm{T},k}^{\max} = P_\mathrm{T}$, and $p_{\mathrm{X},m}^{\max} = P_\mathrm{X}$ for $k \in \mathcal{K}, m \in \mathcal{M}$. The channel realizations are flat in time and frequency. The channel coefficients on the two hops are generated as i.i.d. zero-mean, unit-variance, complex Gaussian random variables. Note that there is no path loss assumed in the simulations, thus all cross-links on the same hop has the same average power. The plots are produced by averaging the solutions for 100 random channel realizations. In each channel realization, the initial transceivers are chosen randomly subject to the power constraints at the transmitters and at the relays. In each iteration, either one transmitter or one relay is allowed to update its precoder, after that all the receivers are allowed to update their

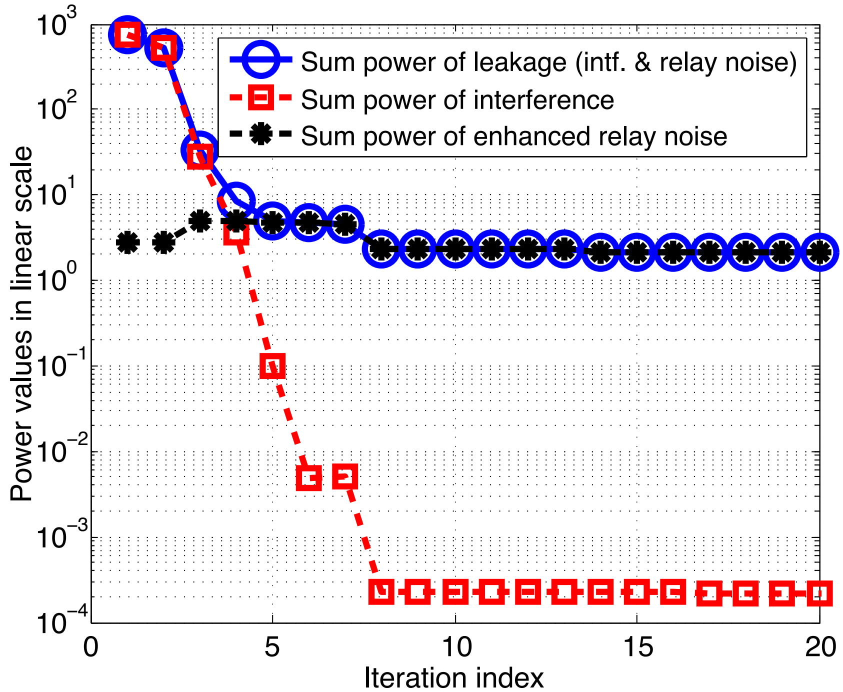

Figure 3.2: The total leakage power at the receivers over iterations of Algorithm 1 for a channel realization of $(4 \times 4, 2)^3 + 4^3$.

receive filters. The CVX toolbox [77] is used to solve convex optimization problems in each iteration.

3.6.1 Convergence

Fig. 3.2 and Fig. 3.3 illustrate the convergence behavior of the proposed algorithms. Fig. 3.2 provides the analysis of the sum power of post-processed leakage signals of Algorithm 1 for a random channel realization of the $(4 \times 4, 2)^3 + 4^3$ system. I observe that the total leakage power decreases monotonically over iterations. Interestingly, interference and enhanced relay

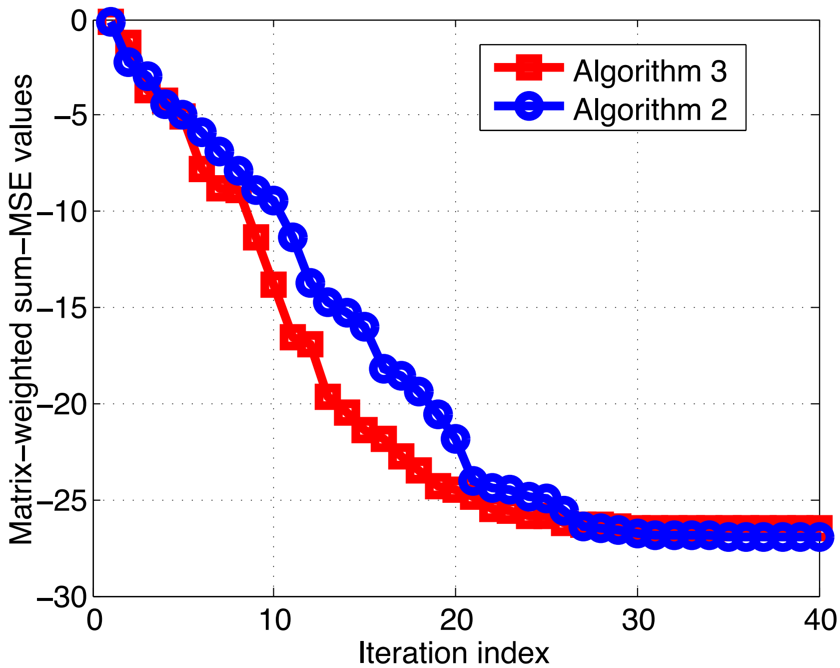

Figure 3.3: Weighted sum-MSE values over iterations of Algorithm 2 and Algorithm 3 for a channel realization of $(2 \times 4, 1)^4 + 2^4$.

noise change their roles during the process of Algorithm 1. The interference is the dominant component of the leakage signal at the beginning, however, it is aligned and then cancelled in a few iterations. After this point, enhanced relay noise becomes dominant - its sum power is orders of magnitude larger than the interference sum power. Unfortunately, given that many space dimensions are devoted to minimizing the sum true interference power, it becomes challenging for Algorithm 1 to align and cancel enhanced relay noise power. Intuitively, enhanced relay noise can be thought of as a source of single-hop interference from "virtual uncoordinated relays" that directly impacts the receivers. Thus, it is necessary to take into account both interference and enhanced re-

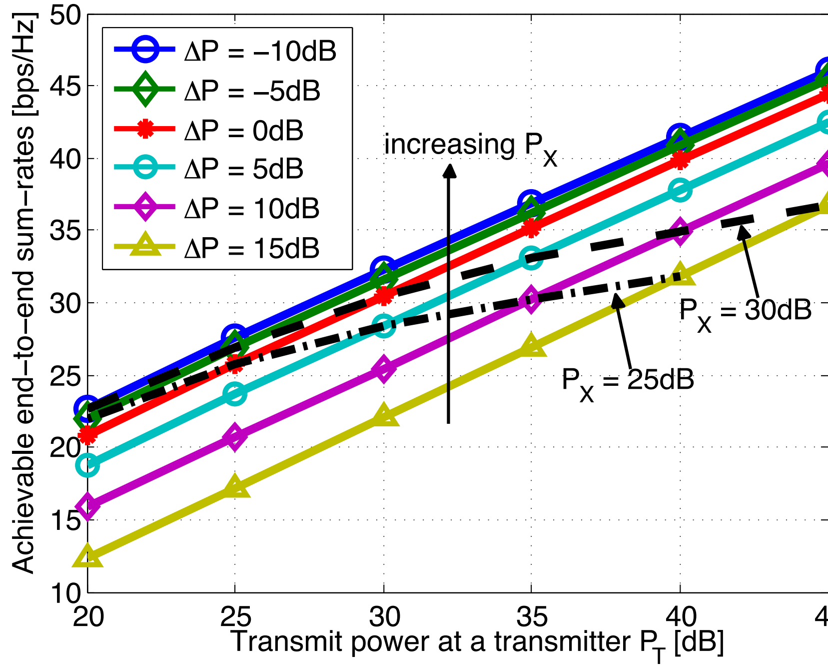

Figure 3.4: Average end-to-end sum-rates of $(4 \times 4, 2)^3 + 4^3$ achieved by Algorithm 3. The solid curves are for different values of ΔP. The dashed line is for $P_{\mathrm{X}} = 30$ dB while the dot-dashed line is for $P_{\mathrm{X}} = 25$ dB.

lay noise in the design of interference alignment strategies for the AF relay interference channel. Fig. 3.3 provides the values of $WMSE_{\text{sum}}$ achieved by Algorithm 2 and by Algorithm 3 over iterations for a channel realization of the $(2 \times 4, 1)^4 + 2^4$ system. I observe that $WMSE_{\text{sum}}$ values for both algorithms are non-increasing over iterations, i.e., the proposed algorithms are convergent. Although the convergence rates are fast for these configurations, it should be noted that the running time increases in both K and d.

3.6.2 Impact of Transmit Power P_{T} and P_{X}

I use the proposed algorithms to obtain insights into the achievable end-to-end sum-rates and multiplexing gains of the AF relay interference channel. The achievable end-to-end multiplexing gains are computed as the slope of the curves of the achievable end-to-end sum-rates at high transmit power values. In this experiment, I investigate the impact of P_{T} and P_{X} on the achievable end-to-end multiplexing gains of Algorithm 3 for a $(4 \times 4, 2)^3 + 4^3$ AF relay system. I denote $\Delta P = P_{\mathrm{T}} - P_{\mathrm{X}}$. Fig. 3.4 presents the achievable end-to-end sum-rates as a function of P_{T} for different values of ΔP, which are represented by the solid lines. I observe that Algorithm 2 achieves a multiplexing gain of 6 for all the simulated values of ΔP. Thus, it is sufficient to consider $\Delta P = 0$ in the following experiments. In addition, I observe that when P_{T} is fixed, the average end-to-end sum-rates is still increasing with P_{X} but at a sublinear rate. A similar observation can be made when P_{X} is fixed while P_{T} is varied. In other words, Algorithm 2 cannot obtain multiplexing gains higher than one if either P_{T} or P_{X} is fixed.

3.6.3 End-to-End Sum-Rate Comparison of the Proposed Algorithms

In this experiment, I compare the average achievable end-to-end sum-rates of the proposed algorithm for the $(2 \times 4, 1)^4 + 2^4$ system as shown in Fig. 3.5. I consider a sum power constraint at the relays and use the same initial condition for all the algorithms. Thanks to power control, Algorithm 3 always outperforms Algorithm 2 in this experiment. Both Algorithm 2 and Algorithm 3 perform significantly better than Algorithm 1 at low-to-medium SNR values. This is because Algorithm 1 does not take into account the desired signal power and the noise power at the receivers, while the others do. Interestingly, at high SNR values, Algorithm 1 outperforms both Algorithm 2 and Algorithm 3 in terms of end-to-end sum-rates and multiplexing gain. Considering the achievable end-to-end rates per user for Algorithm 2 and Algorithm 3, some users have much smaller rates than do the others. Algorithm 3 even turns off the data streams associated with some users. Thus, the solution does not provide fairness among users which limits the maximum end-to-end multiplexing gains achievable by the two algorithms. Thus, Algorithm 1 is more suitable than the others for investigating the maximum achievable end-to-end multiplexing gains of MIMO AF relay networks.

3.6.4 Comparison with DF Relaying and Direct Transmission

To obtain insights into relay functionality selection in the presence of interference, I simulate the dedicated DF relay interference channel where one DF relay is dedicated to assisting one and only one transmitter-receiver pair, i.e., $K = M$. Using equal time-sharing, the end-to-end achievable rate of a pair is defined as half of the minimum between the achievable rate from the transmitter to the associate DF relay and that from the relay to the receiver. Based on [166], I derive an upper-bound on the achievable end-to-end multiplexing gain $d_{\sum}^{\mathrm{DF}}$ of the dedicated DF relay interference channel $(N_{\mathrm{R}} \times N_{\mathrm{T}}, d)^{K} + N_{\mathrm{X}}^{K}$

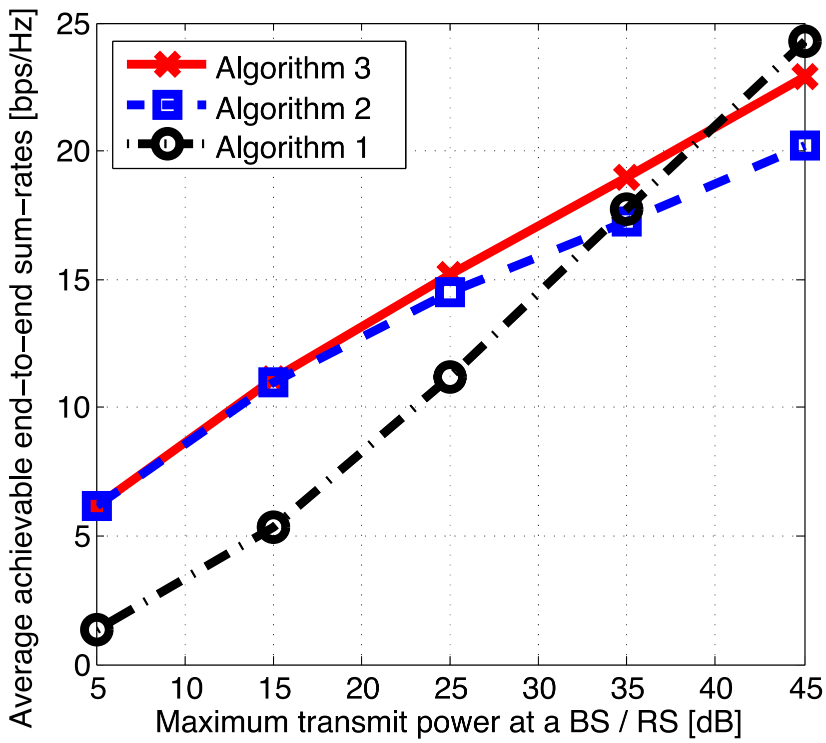

Figure 3.5: Comparison of the average achievable end-to-end sum-rates of the proposed algorithms for the $(2 \times 4, 1)^4 + 2^4$ system.

as follows

$$d_{\Sigma}^{\mathrm{DF}} \leq d_{\Sigma}^{\mathrm{DF,max}} = 0.5 * \min \left\{ \left\lfloor \frac{K(N_{\mathrm{X}} + N_{\mathrm{T}})}{K+1} \right\rfloor, \left\lfloor \frac{K(N_{\mathrm{R}} + N_{\mathrm{X}})}{K+1} \right\rfloor \right\}.$$

I also assume that both the transmitters and the DF relays are subject to individual power constraints with the same average maximum transmit power per node of P_{T} and P_{X}.

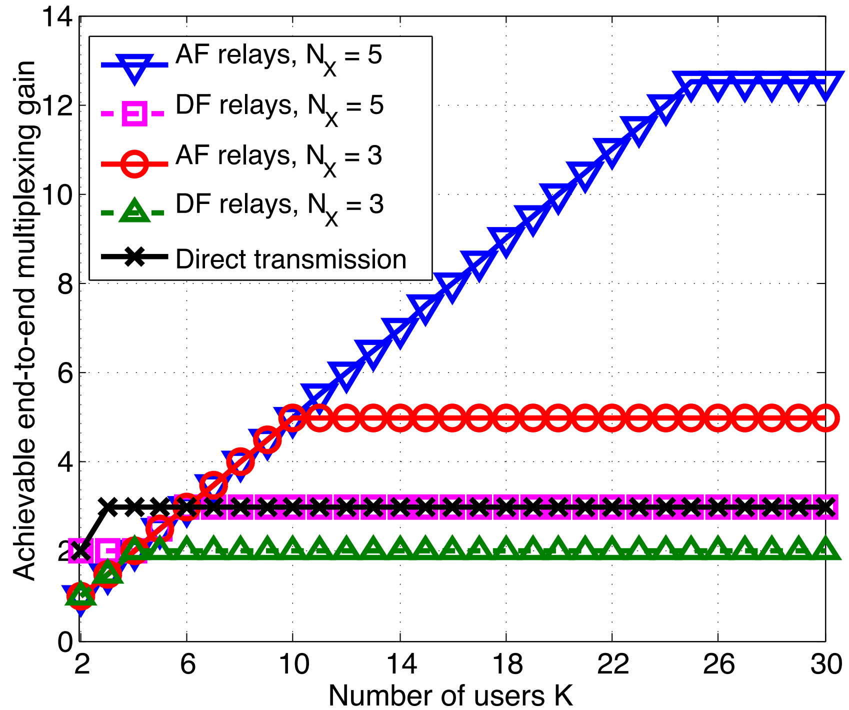

Figure 3.6: Achievable end-to-end multiplexing gains as functions of K for the $(2 \times 2, 1)^K + N_{\mathrm{X}}^K$ systems.

3.6.4.1 Maximum Achievable Multiplexing Gains

In this experiment, Algorithm 1 is used to investigate the maximum achievable end-to-end multiplexing gains for AF relays, DF relays, and direct transmission. Fig. 3.6 shows the results as a function of K for $d = 1$, $N_{\mathrm{R}} = N_{\mathrm{T}} = 2$, and $N_{\mathrm{X}} = 3$ or $N_{\mathrm{X}} = 5$. With these values of N_{X} and when K is small, due to the half-duplex loss, both AF relays and DF relays achieve lower end-to-end multiplexing gains than the direct transmission. While DF relays cannot outperform the direct transmission, AF relays can achieve higher multiplexing gains when there are more than 6 users. Thus, AF relays help increase the achievable end-to-end multiplexing gains of interference channels. In addition, I observe that there exist upper-bounds on the achievable end-to-end multiplexing gains for all the simulated cases - AF relays, DF relays, and direct transmission. Finding closed-form expressions of such bounds in the MIMO AF relay interference channel is left for future work.

3.6.4.2 Achievable End-to-End Sum-Rates

I compare the end-to-end sum-rate performance of Algorithm 2 with other existing transceiver design strategies for the relay interference channel. Specifically, I simulate two strategies for AF relays: i) time division multiple access (TDMA) distributed beamforming (BF) and ii) dedicated relay BF. In the AF TDMA distributed BF, all the relays help only one transmitter-receiver pair at a time (which is an extension of the design in [121] for multiple-antenna receivers). In the dedicated relay BF, I assume that each AF relay is dedicated to assisting one and only one transmitter-receiver pair. This means that interference is ignored and I independently apply the joint source-relay design in [59, 170] for the two-hop channels from the transmitters to their associated receivers. I also simulate different strategies for the MIMO single-hop interference channel on the two hops, including selfish (SF) beamforming (i.e., each transmitter aims at maximizing the achievable rate to its associated receiver), interference alignment based on total leakage (TL) minimization [156], and iteratively weighted MSE sum-rate (SR) maximization strategy [187]. I assume the same strategy is used on two hops.

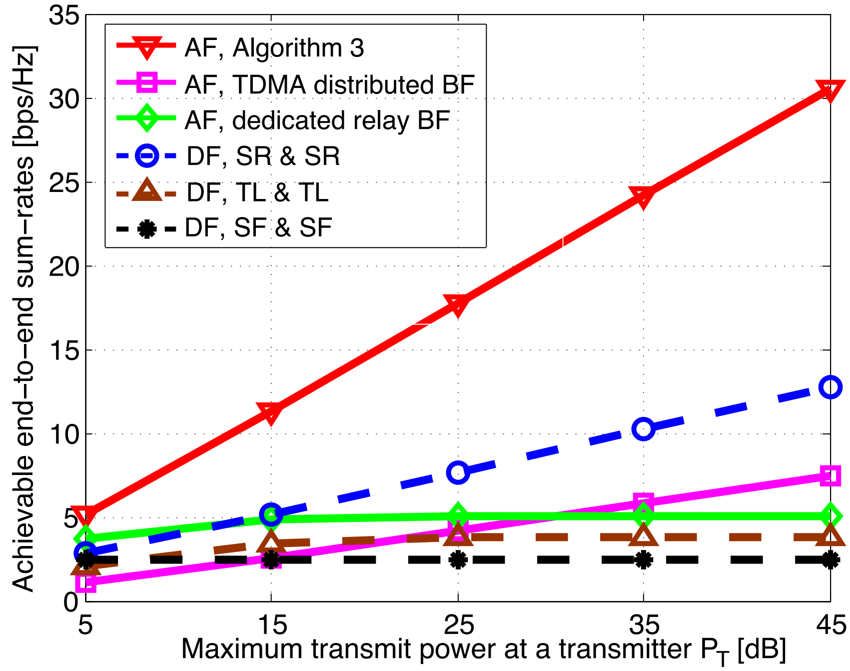

Figure 3.7: Achievable end-to-end sum-rates for the $(2 \times 2, 1)^4 + 2^4$ system.

Fig. 3.7 shows the results for the $(2 \times 2, 1)^4 + 2^4$ system. Note that Algorithm 3 outperforms the other strategies in all regions. It achieves an end-to-end multiplexing gain of 2 (which is equal to half of the total number of data streams). Neglecting interference, the dedicated relay strategies under both AF relays and DF relays achieve zero multiplexing gains. While the end-to-end multiplexing gain achieved by the DF TL & TL strategy is zero, that achieved by the DF SR & SR strategy is nonzero. Because interference alignment is infeasible on two hops, interference cannot be completely eliminated using the TL algorithm. On the contrary, the SR algorithm is able to turn off some data streams, one data stream on each hop in this case, to

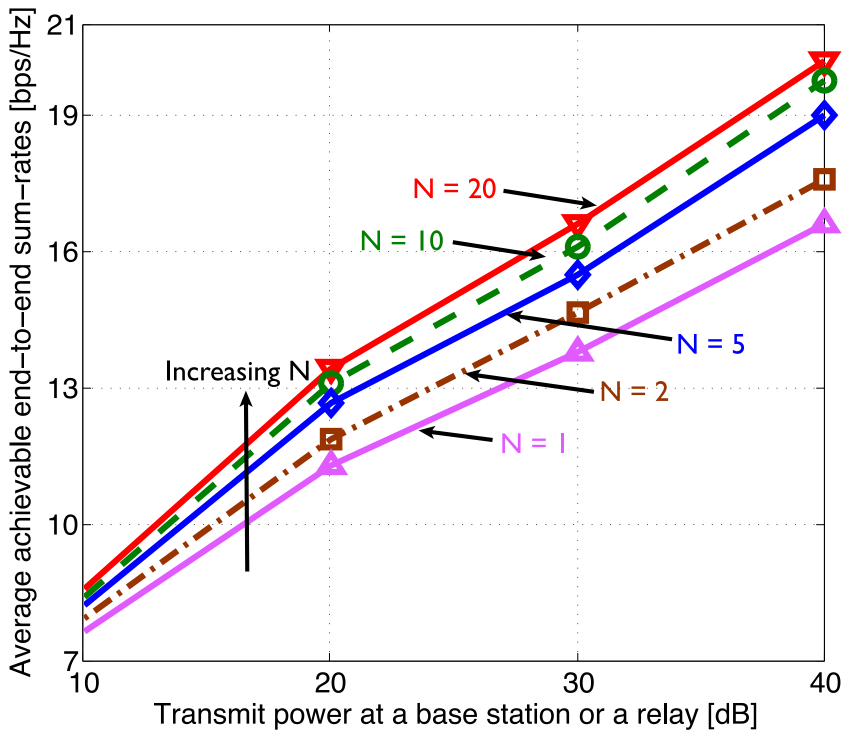

Figure 3.8: Achievable end-to-end sum-rates for the $(2 \times 2, 1)^4 + 2^4$ system with different numbers of initial solutions $N = \{1, 2, 5, 10, 20\}$.

make interference alignment feasible. Note, however, that it may turn off data streams of different pairs on two hops. Thus, on average, the DF SR & SR strategy achieves an end-to-end multiplexing gain less than 1.5 (half of the number of remaining data streams when interference alignment is feasible). Finally, thanks to orthogonalized transmission, the AF TDMA distributed BF can achieve an end-to-end multiplexing gain of 0.5.

3.6.4.3 Opportunistic Approach

The end-to-end sum-rate performance of the stationary points found by Algorithm 2 and Algorithm 3 depend significantly on the initial solutions. The opportunistic approach proposes to use multiple initial solutions and then chooses the one with the highest end-to-end sum-rates. With enough initial solutions, the opportunistic approach should approximate the global solution. Let N denote the number of random initial solutions. Fig. **??** shows the average end-to-end sum-rates for several values of N achieved by Algorithm 3 in the $(2 \times 2, 1)^4 + 2^4$ system. For this setting, at a transmit power of 30dB, the gain provided by the opportunistic approach over the non-opportunistic approach is 6.4% for $N = 2$, 13.2% for $N = 5$, 16.9% for $N = 10$, and 20.6% for $N = 20$. Note that the higher the value of N, the larger the average achievable end-to-end sum-rates. Also, the additional gains obtained by using an extra random initial solution decreases in N. Thus we suspect the gap between our proposed technique and the optimal approach is not too large. Nevertheless, the benefits of this opportunistic approach come at the expense of longer running time.

Chapter 4

Cooperative Transmit Precoding for the MIMO Relay Interference Broadcast Channel

This chapter proposes an algorithm for designing the linear transmit precoders at the transmitters and relays of the relay interference broadcast channel, a generic model for relay-based cellular systems, to maximize the end-to-end sum-rates. Section 4.1 presents the motivations, reviews prior work, and introduces the contributions. Section 4.2 describes the system model. Section 4.3 formulates the design problem and discusses the challenges. Section 4.4 presents the proposed algorithm in detail. Section 4.5 numerically evaluates the achievable end-to-end sum-rates of the proposed algorithm.

4.1 Introduction

Relay communication, a viable solution for coverage extension and capacity enhancement in cellular systems [5, 98], is sensitive to interference. This chapter considers the relay interference broadcast channel where a stage of relays assist multiple transmitters to communicate with multiple receivers. Each receiver intends to receive data from a single transmitter with the aid of a single relay. Multiple relays are dedicated to assisting a single transmitter at the same time; each relay simultaneously forwards data from this transmitter to multiple receivers. Using shared radio resources, the transmissions from the transmitters to relays interfere with each other. Similarly, the transmissions from the relays to receiver interfere with each other. If considered separately, each hop of this model is an instance of the single-hop interference broadcast channel [165, 187], a generalization of the conventional single-hop interference channel. Each transmitter in the broadcast channel has data for multiple receivers while each transmitter in the conventional channel has data for only one receiver. Although recent results on the single-hop interference broadcast channel [165, 187] can be applied separately for the two hops, even higher end-to-end sum-rates can be achieved if the transmitters and relays are configured jointly. Unfortunately, jointly configuring the transmitters and relays is challenging, especially with limited information about the interferers.

In this chapter, I focus on half-duplex DF relays that cannot transmit and receive at the same time. By assumption, relays decode only signals intended to their associated receivers. I assume all two-hop links have a common timesharing value, i.e., the same fractions of time for transmission on two hops. The achievable end-to-end rate corresponding to a relay is defined as the minimum of the achievable normalized rate from its associated transmitter to itself and the achievable normalized sum-rates from itself to its associated receivers [36]. A two-hop rate mismatch occurs when some links have a dominant first hop while others have a dominant second hop, resulting in unused rates on two hops and hence low end-to-end sum-rates. An efficient system design should not cause any two-hop rate mismatch while mitigating interference.

In general, the optimal transmit and receive strategies for sum-rate maximization in interference channels are not widely known, even for the single-hop one. Thus, I adopt a pragmatic approach that treats interference as noise and maximizes end-to-end sum-rates by searching within the class of linear transmit and receive strategies. Assuming the receivers always use the optimal linear MMSE receive filters, I focus on designing the transmit precoders at the transmitters and relays. Unfortunately, the problem is non-convex and NP-hard. Moreover, it falls into a class of max-min optimization problems. Thus, finding the stationary points of the problem, including its globally and locally optimal solutions, is challenging [105, 110, 123].

Transmit precoder design has been studied widely for the multiple-antenna single-hop interference broadcast channel [165, 187], especially for its special case of the single-hop interference channel [70, 88, 94, 105, 115, 158, 164, 177, 183, 184, 186, 228]. The methods in [94, 183, 184, 186] were based on the so-called interference pricing framework where the transmitters configure themselves based on interference prices fed back from the receivers. Interference prices represent the marginal decrease in the sum-utility function per unit increase in interference power. In [165, 177, 187], based on a relationship between mutual information and mean squared errors, the authors proposed to solve sum-utility maximization problems via iterative minimization of weighted sum-MSEs. Under certain conditions on the utility functions, the algorithms in [94, 165, 177, 183, 184, 186, 187] are guaranteed to converge to the stationary points of the corresponding sum-utility maximization problems. Note that the existing single-hop results are designed specifically for the single-hop interference channel. Thus, it is not straightforward how to extend them to take into special features of the relay interference channel, like relay signal processing operation and multi-hop transmission.

To the best of my knowledge, there has been little prior work on transmit precoder design in the relay interference broadcast channel. Much prior work, however, focused on interference mitigation for special cases of this model. In [35], the authors proposed a transmit precoder design for the MIMO relay broadcast channel where a single MIMO DF relay forwards data from a single transmitter to multiple receivers. This means that there is no interference on the first hop and there is no inter-relay interference on the second hop. My prior work in [208] considered the DF relay interference channel where the receivers are equipped with a single antenna and each relay is dedicated to aiding a single transmitter-receiver pair. Based on the interference pricing framework, the algorithm in [208] used approximations of end-to-end rates to compute interference prices for designing the second-hop transmit precoders with fixed first-hop transmit precoders. Nevertheless, it is not straightforward to extend it to the general relay interference broadcast channel with multiple-antenna receivers. There have been many other algorithms for designing transmit precoders at the relays and/or transmitters in the relay interference channel [7, 38, 40, 60, 72, 106, 122, 144–147, 150, 159, 163, 232]. Nevertheless, they considered either AF relays [7, 38, 40, 60, 72, 122, 144–147, 163, 232] or other relay architectures, like the shared relays [150, 159] or two-way relays [106]. In addition, in this prior work, each relay simultaneously forwards data for multiple transmitter-receiver pairs unlike in my approach.

In this chapter, I propose a cooperative algorithm for efficiently finding suboptimal solutions of the transmit precoder design problem with high end-to-end sum-rates. The proposed algorithm can be implemented in a distributed fashion with low communication overhead. The proposed algorithm consists of three phases in the following order: i) second-hop transmit precoder design, ii) first-hop transmit precoder design, and iii) first-hop transmit power control. In the first phase, I propose to ignore the first hop and focus on configuring the relays to maximize the achievable second-hop sum-rates. Essentially, the second hop is treated like the conventional single-hop interference broadcast channel. Thus, existing single-hop algorithms can be applied to find the stationary points of second-hop sum-rate maximization [70, 88, 94, 105, 115, 158, 164, 165, 177, 183, 184, 186, 187, 228]. Having computed the second-hop transmit precoders, each relay computes the sum of achievable rates from itself to its associated receivers, which is used as input to the second phase.

The second phase focuses on designing the first-hop transmit precoders. In the naïve approach, this can be done by applying the prior work in [70, 88, 94, 105, 115, 158, 164, 165, 177, 183, 184, 186, 187, 228] while ignoring the designed second-hop transmit precoders. The naïve approach, however, may cause a two-hop rate mismatch because it cannot take into account the time-sharing value and second-hop configuration. To overcome this limitation, I propose to formulate and solve a new problem to maximize an approximation of the achievable end-to-end sum-rates. Such approximations of the achievable end-to-end rates depend not only on the first-hop transmit precoders, but also on the timesharing value and second-hop configuration. This allows for second-hop interference mitigation at the same time as rate-mismatch alleviation. Some guidelines for selecting such approximations are provided. Having defined a more comprehensive utility function, I use the technique in [165, 177, 187] to develop an iterative method that is guaranteed to converge to the stationary points of the new sum-utility maximization problem. This concludes the second phase of the proposed algorithm.

The output of the second phase, however, may still contain a residual two-hop rate mismatch since only an approximate solution is proposed. Consequently, in the final phase, I propose to fix the shapes of the first-hop transmit precoders and to adjust their norms to eliminate completely two-hop rate mismatching. Essentially, this is a transmit power control problem. Note that for a two-hop link with a dominant first hop, excess power is allocated for the transmissions on the first hop. I propose a method for simultaneously reducing excess power for the first-hop transmissions so that the achievable end-to-end rates for all the relays are nondecreasing over iterations, thus potentially improving the achievable end-to-end sum-rates. The method is guarantee to converge. At the convergence point, there are no two-hop links with a dominant first hop, i.e., there is not any two-hop rate mismatch. Although there have been many power control algorithms for the single-hop interference channel [63, 94, 160, 175, 185, 193–195, 197, 224] and for the relay interference channel [188, 189, 218, 235], they are not designed to eliminate two-hop rate mismatching. Therefore, even if applicable to the third phase, existing power control algorithms may worsen the two-hop rate mismatch situation.

I use Monte Carlo simulations to evaluate the average achievable end-to-end sum-rates of the proposed algorithm for various relay interference broadcast channel configurations. The naïve approach of applying existing single-hop results separately for the two hops is selected as the baseline strategy. The proposed algorithm and the naïve approach have the same second-hop transmit precoders. While the timesharing value and second-hop configuration are taken into account in the first-hop transmit precoder design in the last two phases of the proposed algorithm, they are ignored in that of the naïve approach. Numerical results show that the first two phases of the proposed algorithm are enough to provide large end-to-end sum-rate gains over the naïve approach. In addition, the last phase of the proposed algorithm makes considerable improvements in end-to-end sum-rate performance over the output of the second phase. This highlights the importance of two-hop rate matching in the DF relay interference (broadcast) channel. Finally, each phase of the proposed algorithm converges in a few iterations. Note that the proposed algorithm can be implemented in a distributed manner with a little more overhead than the naïve approach. Thus, it is suitable for practical implementation in DF relay networks.

4.2 System Model

Consider a relay interference broadcast channel where K_{T} transmitters communicate with K_{R} receivers with the aid of K_{X} half-duplex decode-and-forward relays, as illustrated in Fig. 4.1. Each transmitter is assigned a unique index from $\mathcal{K}_{\mathrm{T}} \triangleq \{1, \cdots, K_{\mathrm{T}}\}$. Similarly, each relay is assigned a unique index from $\mathcal{K}_{\mathrm{X}} \triangleq \{1, \cdots, K_{\mathrm{X}}\}$ and each receiver is assigned a unique index from $\mathcal{K}_{\mathrm{R}} \triangleq \{1, \cdots, K_{\mathrm{R}}\}$. Each transmitter may require the aid of multiple relays to simultaneously send independent data streams to its receivers. Each relay is dedicated to serving multiple receivers associated with a single transmitter. Each receiver intends to receive data from only one transmitter with the aid of a single relay. Let $\chi(k) \in \mathcal{K}_{\mathrm{X}}$ denote the index of the relay that aids receiver $k \in \mathcal{K}_{\mathrm{R}}$. Let $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ denote the index of the transmitter that is aided by relay $k \in \mathcal{K}_{\mathrm{X}}$. The transmitters and relays do not share data. I assume that each relay k does not attempt to decode the signal intended for receiver $m \in \mathcal{K}_{\mathrm{R}}$ with $\chi(m) \neq k$. Transmitter $k \in \mathcal{K}_{\mathrm{T}}$ has $N_{\mathrm{T},k}$ antennas, relay $m \in \mathcal{K}_{\mathrm{X}}$ has $N_{\mathrm{X},m}$ antennas, and receiver $q \in \mathcal{K}_{\mathrm{R}}$ has $N_{\mathrm{R},q}$ antennas.

Half-duplex relays cannot transmit and receive at the same time, thus the transmission procedure requires two stages. Using a common frequency at the same time, the transmissions in the same stage interfere with each other. For tractable analysis, I assume Gaussian signaling is used in both stages although it may not be optimal for the relay interference broadcast channel. Specifically, the transmitters, including uncoordinated interferers, use independent Gaussian codebooks for transmission. This allows the relays and

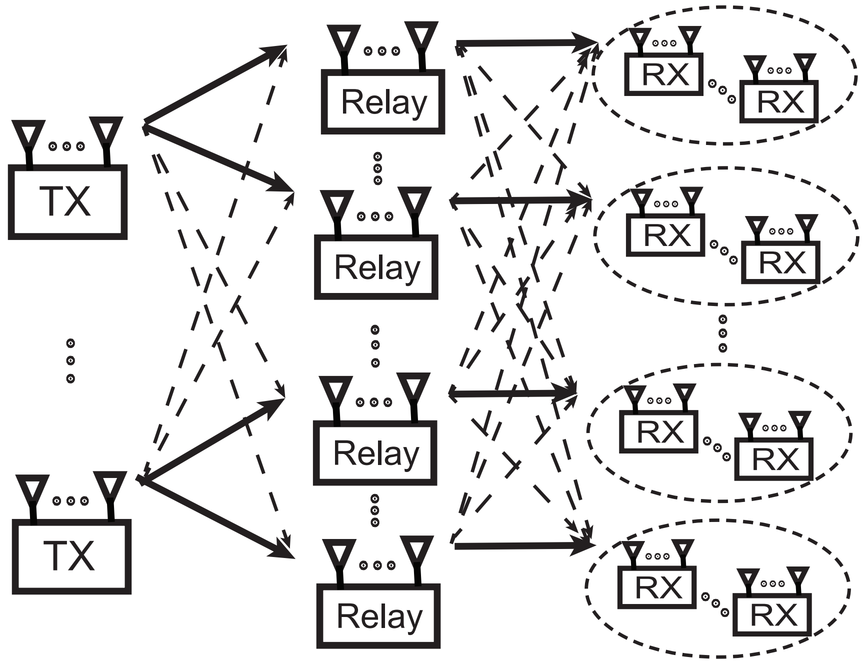

Figure 4.1: A relay interference broadcast channel where a number of half-duplex decode-and-forward relays aid the data transmission from a number of transmitters to their associated receivers. The solid lines connect the associated nodes and represent communication links. The dashed lines represent interference links.

receivers to treat interference as additive Gaussian noise while decoding their desired signals. In the first stage, the transmitters send data to the relays. Treating unwanted signals as additive Gaussian noise, each relay decodes its desired signal. Each relay separates its decoded signals for individual associated receivers, re-encodes and retransmits to its associated receivers in the second stage. Each receiver also treats unwanted signals as additive Gaussian noise when decoding its desired signal. Similar to chapter 3, I assume that nonGaussian interference and noise are out of the scope of this chapter.

I assume that the direct channels are neglected by the second-stage receivers. I also assume the relay interference channel is fully-connected on each hop. Specifically, each relay receives non-negligible signals from all the transmitters and each receiver observes non-negligible signals from all the relays. This assumption is good for modeling certain scenarios in cellular networks. For example, it can model several adjacent cells in a dense cellular network. Relays are deployed at good geographical locations, e.g., with high altitudes and clear paths to the base stations, to aid the base stations communicate with the users. Several relays may be used to aid the communication between one base station and a number of users located near the edge of its cells. Due to its high altitude and low-shadowed paths to the base stations, each relay observes non-negligible unwanted signals from unintended base stations. Similarly, each user observes non-negligible unwanted signals from unintended relays.

I consider slowly-varying, frequency-flat, block-fading channels. I denote $\mathbf{H}_{m,k} \in \mathbb{C}^{N_{\mathrm{X},m} \times N_{\mathrm{T},k}}$ as the matrix channel between transmitter k and relay m for $k \in \mathcal{K}_{\mathrm{T}}$ and $m \in \mathcal{K}_{\mathrm{X}}$. Let $\mathbf{x}_{1,k} \in \mathbb{C}^{d_{1,k} \times 1}$ denote the symbol vector that transmitter $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ sends to relay $k \in \mathcal{K}_{\mathrm{X}}$ in the first stage, where $d_{1,k}$ is the number of data streams and $\mathbb{E}(\mathbf{x}_{1,k}\mathbf{x}_{1,k}^*) = \mathbf{I}_{d_{1,k}}$. I denote $\mathbf{F}_{\mathrm{T},k} \in \mathbb{C}^{N_{\mathrm{T},\mu(k)} \times d_{1,k}}$ as the linear transmit precoder that transmitter $\mu(k)$ uses to send $\mathbf{x}_{1,k}$ to relay $k \in \mathcal{K}_{\mathrm{X}}$. I define $\mathbf{F}_{\mathrm{T}} \triangleq \{\mathbf{F}_{\mathrm{T},1}, \cdots, \mathbf{F}_{\mathrm{T},K_{\mathrm{X}}}\} \in \mathbb{F}_{\mathrm{T}} \triangleq \mathbb{C}^{N_{\mathrm{T},\mu(1)} \times d_{1,1}} \times \cdots \times \mathbb{C}^{N_{\mathrm{T},\mu(K_{\mathrm{X}})} \times d_{1,K_{\mathrm{X}}}}$. Let $p_{\mathrm{T},k}$ be the sum transmit power at transmitter $k \in \mathcal{K}_{\mathrm{T}}$. The sum transmit power constraint at transmitter

$k \in \mathcal{K}_{\mathrm{T}}$ is

$$\begin{aligned} p_{\mathrm{T},k} &\geq \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m)=k}} \|\mathbf{F}_{\mathrm{T},m}\|_{\mathrm{F}}^2 \\ &= \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^*), \forall k \in \mathcal{K}_{\mathrm{T}}. \end{aligned} \tag{4.1}$$

Let $\mathbf{n}_{\mathrm{X},k} \in \mathbb{C}^{N_{\mathrm{X},k} \times 1}$ be the spatially white, additive Gaussian noise at relay $k \in \mathcal{K}_{\mathrm{X}}$ with $\mathbb{E}(\mathbf{n}_{\mathrm{X},k}\mathbf{n}_{\mathrm{X},k}^*) = \sigma_{\mathrm{X},k}^2 \mathbf{I}_{N_{\mathrm{X},k}}$. Relay $k \in \mathcal{K}_{\mathrm{X}}$ observes

$$\mathbf{y}_{\mathrm{X},k} = \underbrace{\mathbf{H}_{k,\mu(k)}\mathbf{F}_{\mathrm{T},k}\mathbf{x}_{1,k}}_{\text{desired signal}} + \underbrace{\sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ m \neq k}} \mathbf{H}_{k,\mu(m)}\mathbf{F}_{\mathrm{T},m}\mathbf{x}_{1,m}}_{\text{interference}} + \mathbf{n}_{\mathrm{X},k}. \tag{4.2}$$

The interference plus noise covariance matrix at relay $k \in \mathcal{K}_{\mathrm{X}}$ is

$$\mathbf{R}_{\mathrm{X},k}(\mathbf{F}_{\mathrm{T}}) = \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ m \neq k}} \mathbf{H}_{k,\mu(m)}\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^*\mathbf{H}_{k,\mu(m)}^* + \sigma_{\mathrm{X},k}^2 \mathbf{I}_{N_{\mathrm{X},k}}. \tag{4.3}$$

Each relay $k \in \mathcal{K}_{\mathrm{X}}$ applies a linear receive filter $\mathbf{W}_{\mathrm{X},k} \in \mathbb{C}^{N_{\mathrm{X},k} \times d_{1,k}}$ to $\mathbf{y}_{\mathrm{X},k}$. I define $\mathbf{W}_{\mathrm{X}} \triangleq (\mathbf{W}_{\mathrm{X},1}, \cdots, \mathbf{W}_{\mathrm{X},K_{\mathrm{X}}}) \in \mathbb{W}_{\mathrm{X}} \triangleq \mathbb{C}^{N_{\mathrm{X},1} \times d_{1,1}} \times \cdots \times \mathbb{C}^{N_{\mathrm{X},K_{\mathrm{X}}} \times d_{1,K_{\mathrm{X}}}}$. The maximum achievable rate from transmitter $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ to relay $k \in \mathcal{K}_{\mathrm{X}}$ is obtained by the following linear MMSE receive filter

$$\mathbf{W}_{\mathrm{X},k}^{\mathrm{MMSE}} = \left[\mathbf{H}_{k,\mu(k)}\mathbf{F}_{\mathrm{T},k}\mathbf{F}_{\mathrm{T},k}^*\mathbf{H}_{k,\mu(k)}^* + \mathbf{R}_{\mathrm{X},k}(\mathbf{F}_{\mathrm{T}})\right]^{-1} \mathbf{H}_{k,\mu(k)}\mathbf{F}_{\mathrm{T},k}. \tag{4.4}$$

Thus, the maximum achievable rate on the first hop at relay $k \in \mathcal{K}_{\mathrm{X}}$ is

$$R_{1,k}(\mathbf{F}_{\mathrm{T}}) = \log_2 \det \left(\mathbf{I} + \mathbf{F}_{\mathrm{T},k}^*\mathbf{H}_{k,\mu(k)}^* \left[\mathbf{R}_{\mathrm{X},k}(\mathbf{F}_{\mathrm{T}})\right]^{-1} \mathbf{H}_{k,\mu(k)}\mathbf{F}_{\mathrm{T},k}\right) \tag{4.5}$$

$$= \log_2(1 + \xi_{1,k}(\mathbf{F}_{\mathrm{T}})), \tag{4.6}$$

where $\xi_{1,k}(\mathbf{F}_{\mathrm{T}})$ is the effective first-hop SINR corresponding to relay k.

Let $\mathbf{G}_{m,k} \in \mathbb{C}^{N_{\mathrm{R},m} \times N_{\mathrm{X},k}}$ denote the matrix channel between relay $k \in \mathcal{K}_{\mathrm{X}}$ and receiver $m \in \mathcal{K}_{\mathrm{R}}$. Let $\mathbf{x}_{2,k} \in \mathbb{C}^{N_{\mathrm{X},k} \times d_{2,k}}$ denote the transmit symbol vector that relay $\chi(k) \in \mathcal{K}_{\mathrm{X}}$ sends to receiver $k \in \mathcal{K}_{\mathrm{R}}$ where $d_{2,k}$ is the number of data streams and $\mathbb{E}(\mathbf{x}_{2,k}\mathbf{x}_{2,k}^*) = \mathbf{I}_{d_{2,k}}$. I denote $\mathbf{F}_{\mathrm{X},k} \in \mathbb{C}^{N_{\mathrm{X},\chi(k)} \times d_{2,k}}$ as the linear transmit precoder that relay $\chi(k)$ uses to send $\mathbf{x}_{2,k}$ to receiver $k \in \mathcal{K}_{\mathrm{R}}$. I define $\mathbf{F}_{\mathrm{X}} \triangleq \{\mathbf{F}_{\mathrm{X},1}, \cdots, \mathbf{F}_{\mathrm{X},K_{\mathrm{R}}}\} \in \mathbb{F}_{\mathrm{X}} \triangleq \mathbb{C}^{N_{\mathrm{X},\chi(1)} \times d_{2,1}} \times \cdots \times \mathbb{C}^{N_{\mathrm{X},\chi(K_{\mathrm{R}})} \times d_{1,K_{\mathrm{R}}}}$. Let $p_{\mathrm{X},k}$ be the sum transmit power at relay $k \in \mathcal{K}_{\mathrm{X}}$. The sum transmit power constraint at relay $k \in \mathcal{K}_{\mathrm{X}}$ is as follows

$$\sum_{\substack{m \in \mathcal{K}_{\mathrm{R}} \\ \chi(m)=k}} \|\mathbf{F}_{\mathrm{X},m}\|_{\mathrm{F}}^2 = \sum_{\substack{m \in \mathcal{K}_{\mathrm{R}} \\ \chi(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{X},m}\mathbf{F}_{\mathrm{X},m}^*) \le p_{\mathrm{X},k}, \forall k \in \mathcal{K}_{\mathrm{X}}. \tag{4.7}$$

Let $\mathbf{n}_{\mathrm{R},k} \in \mathbb{C}^{N_{\mathrm{R},k} \times 1}$ be the spatially white, additive Gaussian noise at receiver $k \in \mathcal{K}_{\mathrm{R}}$ with $\mathbb{E}(\mathbf{n}_{\mathrm{R},k}\mathbf{n}_{\mathrm{R},k}^*) = \sigma_{\mathrm{R},k}^2 \mathbf{I}_{N_{\mathrm{R},k}}$. Receiver $m \in \mathcal{K}_{\mathrm{R}}$ observes

$$\mathbf{y}_{\mathrm{R},k} = \underbrace{\mathbf{G}_{k,\chi(k)}\mathbf{F}_{\mathrm{X},k}\mathbf{x}_{2,k}}_{\text{desired signal}} + \underbrace{\sum_{\substack{m \in \mathcal{K}_{\mathrm{R}} \\ m \neq k}} \mathbf{G}_{k,\chi(m)}\mathbf{F}_{\mathrm{X},m}\mathbf{x}_{2,m}}_{\text{interference}} + \mathbf{n}_{\mathrm{R},k}. \tag{4.8}$$

The interference plus noise covariance matrix at receiver $k \in \mathcal{K}_{\mathrm{R}}$ is given by

$$\mathbf{R}_{\mathrm{R},k}(\mathbf{F}_{\mathrm{X}}) = \sum_{\substack{m \in \mathcal{K}_{\mathrm{R}} \\ m \neq k}} \mathbf{G}_{k,\chi(m)}\mathbf{F}_{\mathrm{X},m}\mathbf{F}_{\mathrm{X},m}^*\mathbf{G}_{k,\chi(m)}^* + \sigma_{\mathrm{R},k}^2 \mathbf{I}_{N_{\mathrm{R},k}}. \tag{4.9}$$

Each receiver $k \in \mathcal{K}_{\mathrm{R}}$ applies a linear receive filter $\mathbf{W}_{\mathrm{R},k} \in \mathbb{C}^{N_{\mathrm{R},k} \times d_{2,k}}$ to $\mathbf{y}_{\mathrm{R},k}$. The maximum achievable rate from relay $\chi(k) \in \mathcal{K}_{\mathrm{X}}$ to receiver $k \in \mathcal{K}_{\mathrm{R}}$ is achieved by the following linear MMSE receive filter

$$\mathbf{W}_{\mathrm{R},k}^{\mathrm{MMSE}} = \left[\mathbf{G}_{k,\chi(k)}\mathbf{F}_{\mathrm{X},k}\mathbf{F}_{\mathrm{X},k}^*\mathbf{G}_{k,\chi(k)}^* + \mathbf{R}_{\mathrm{R},k}(\mathbf{F}_{\mathrm{X}})\right]^{-1} \mathbf{G}_{k,\chi(k)}\mathbf{F}_{\mathrm{X},k}. \tag{4.10}$$

Thus, the maximum achievable rate at receiver $k \in \mathcal{K}_{\mathrm{R}}$ is

$$R_{2,k}(\mathbf{F}_{\mathrm{X}}) = \log_2 \det \left(\mathbf{I} + \mathbf{F}_{\mathrm{X},k}^{*} \mathbf{G}_{k,\chi(k)}^{*} \left[\mathbf{R}_{\mathrm{R},k}(\mathbf{F}_{\mathrm{X}})\right]^{-1} \mathbf{G}_{k,\chi(k)} \mathbf{F}_{\mathrm{X},k}\right). \quad (4.11)$$

The sum of maximum achievable second-hop rates for relay $k \in \mathcal{K}_{\mathrm{X}}$ is defined as

$$\begin{aligned} R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}}) &\triangleq \sum_{\substack{m \in \mathcal{K}_{\mathrm{R}} \\ \chi(m)=k}} R_{2,m}(\mathbf{F}_{\mathrm{X}}) &\quad (4.12)\\ &= \log_2(1 + \xi_{2,k}(\mathbf{F}_{\mathrm{X}})), &\quad (4.13) \end{aligned}$$

where $\xi_{2,k}(\mathbf{F}_{\mathrm{X}})$ is the effective second-hop SINR corresponding to relay k.

I assume the transmissions in each stage start and end at the same time. Let t be the fraction of time for transmission on the first hop, which is also referred to as the timesharing value. The fraction of time for transmission on the second hop is $(1 - t)$. For example, in 3GPP LTE-Advanced cellular systems, t depends on the number of subframes for backhaul links (i.e., between base stations and relays) in a radio frame [5]. For relay $k \in \mathcal{K}_{\mathrm{X}}$, the normalized achievable rate on the first hop is $tR_{1,k}(\mathbf{F}_{\mathrm{T}})$ while the sum of achievable second-hop rates for this relay is $(1 - t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$. Based on the relative comparison of the normalized rates on two hops, the two-hop link corresponding to relay $k \in \mathcal{K}_{\mathrm{X}}$ can be classified into the following three categories: i) first-hop dominant if $tR_{1,k}(\mathbf{F}_{\mathrm{T}}) > (1 - t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$, ii) second-hop dominant if $tR_{1,k}(\mathbf{F}_{\mathrm{T}}) < (1 - t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$, and iii) equal rate if $tR_{1,k}(\mathbf{F}_{\mathrm{T}}) = (1 - t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$.

The achievable end-to-end rate for relay $k \in \mathcal{K}_{\mathrm{X}}$ is defined as the minimum of the normalized achievable rates on two hops and hence is given by [36]

$$\begin{aligned} R_k(\mathbf{F}_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}}) &\triangleq \min\{tR_{1,k}(\mathbf{F}_{\mathrm{T}}), (1-t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})\} \\ &= \min\{t\log_2(1+\xi_{1,k}(\mathbf{F}_{\mathrm{T}})), (1-t)\log_2(1+\xi_{2,k}(\mathbf{F}_{\mathrm{R}}))\}. \end{aligned} \tag{4.14}$$

The maximum end-to-end sum-rates is defined as

$$R_{\mathrm{sum}}(\mathbf{F}_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}}) \triangleq \sum_{k \in \mathcal{K}_{\mathrm{X}}} R_k(\mathbf{F}_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}}). \tag{4.15}$$

Note that the design of $\mathbf{F}_{\mathrm{T}}$ and $\mathbf{F}_{\mathrm{X}}$ should take into account t.

Remark 4.2.1. This remarks summarizes the key assumptions in this chapter and their justification.

- *Assumption 4.1: A stage of multiple-antenna DF relays aid the transmission from the transmitters to their associated receivers.*
- *Assumption 4.2: The relays cannot transmit and receive at the same time.* This means I consider only half-duplex relays since they are more practical than full-duplex relays.
- *Assumption 4.3: Each receiver is served by only one transmitter via the aid of a single relay. Each relay may forward data from a transmitter to multiple associated receivers. Each transmitter may require the aid of multiple relays.* For example, for the downlink in cellular networks, there may be many users (i.e., receivers) in a cell served by a relay (of which the radius is about 250 meters). Similarly, there may be many relays in a cell served by a base station (of which the radius is a few kilometers).

- *Assumption 4.4: The transmissions on each hop start at the same time and end at the same time.* This can be done based on the use of a common frame structure. This assumption constraints the types of interference at any time, making analysis more tractable.

- *Assumption 4.5: The channels are frequency-flat, slowly-varying, and block-fading.* For example, the results can be used for a single carrier of MIMO OFDM AF relay interference systems.

- *Assumption 4.6: Direct channels from the transmitters and receivers are ignored by the second-stage receivers.* This assumption helps simplify the analysis. Moreover, it is unclear how to take the advantage of the signals sent directly from the transmitters to the receivers in the relay interference channel. This means that without appropriate signal processing at the receivers, signals received directly from the transmitters may even degrade the system performance.

- *Assumption 4.7: Each transmitter has perfect and instantaneous information of the channels from itself to all the relays. Each relay has perfect and instantaneous information of the channels from itself to all the receivers.* Although this is a strict requirement, my results are still valuable since they show the substantial gains that can be achieved through coordination. My results can be used as a benchmark for future work that makes a more practical CSI assumption. Note that I do not assume global CSI in this chapter.

- *Assumption 4.8: Received signals at the relay and receiver are corrupted by additive, circularly symmetric, complex spatially white Gaussian noise. The transmitters, including uncoordinated interferers, use independent Gaussian codebooks for transmission.* This allows the receivers to treat interference as additive Gaussian noise while decoding their desired signals. This assumption is crucial to deriving the rate expressions used in this chapter. NonGaussian interference is not considered in this chapter.

- *Assumption 4.9: Transmitters use linear transmit precoders and receivers use linear receive filters.* The linear processing at the transmitters and relays are attractive in practice due to its low implementation complexity.

- *Assumption 4.10: The relay interference channel is fully-connected on each hop.* For example, this is reasonable for modeling a number of adjacent cells in a dense cellular network.

- *Assumption 4.11: The numbers of data streams and the numbers of antennas satisfy the feasibility conditions of interference alignment on each hop.* This means that interference alignment is always feasible for the MIMO interference broadcast channels on two hops.

4.3 Problem Formulation

This section formulates the design problem and discusses the challenges in finding optimal solutions. The problem for designing the precoders at the transmitters and relays to maximize the sum of achievable end-to-end rates is formulated as follows

$$
\begin{aligned}
(\mathcal{OP}): \quad & \max_{(\mathbf{F}_{\mathrm{T}},\mathbf{F}_{\mathrm{X}})\in\mathbb{F}_{\mathrm{T}}\times\mathbb{F}_{\mathrm{X}}} && R_{\mathrm{sum}}(\mathbf{F}_{\mathrm{T}},\mathbf{F}_{\mathrm{X}}) \\
& \text{subject to} && \sum_{\substack{m\in\mathcal{K}_{\mathrm{R}}\\ \chi(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{X},m}\mathbf{F}_{\mathrm{X},m}^{*}) \le p_{\mathrm{X},k}, \forall k\in\mathcal{K}_{\mathrm{X}}, \\
& && \sum_{\substack{m\in\mathcal{K}_{\mathrm{X}}\\ \mu(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^{*}) \le p_{\mathrm{T},k}, \forall k\in\mathcal{K}_{\mathrm{T}}.
\end{aligned}
$$

The transmit precoder design problem for sum-rate maximization in the single-hop interference channel is nonconvex and NP-hard [105, 123]. This means that its globally optimal solutions cannot be found efficiently in terms of computational complexity even in a centralized fashion. The more complicated problem, $(\mathcal{OP})$ is expected to be NP-hard as well. Moreover, since $(\mathcal{OP})$ is a complicated max-min problem, it is challenging to find the stationary points of $(\mathcal{OP})$, including its optimal solutions [110]. In this chapter, my main focus is to find suboptimal solutions to $(\mathcal{OP})$ with high values of achievable end-to-end sum-rates.

Two main challenges in solving for high-quality suboptimal solutions of $(\mathcal{OP})$ are presented in Remark 4.3.1 and Remark 4.3.2.

Remark 4.3.1. Interference mitigation is a key challenge in end-to-end sum-rate maximization. According to (4.2), each relay observes undesired signals from unintended transmitters on the first hop. Similarly, according to (4.8), each receiver observes undesired signals from unintended relays as well as from its associated relay (but they are intended for other receivers). Due to interference, there exists coupling among the achievable rates on the same hop.

Remark 4.3.2. Two-hop rate matching is another main challenge in maximizing the end-to-end sum-rates of the DF relay interference broadcast channel. Specifically, for a given t, $\mathbf{F}_{\mathrm{T}}$, and $\mathbf{F}_{\mathrm{R}}$, there may exist a mismatch between the normalized achievable rates on two hops. By definition, a two-hop rate mismatch occurs when there exist $k, m \in \mathcal{K}_{\mathrm{X}}$ and $k \neq m$, such that $tR_{1,k}(\mathbf{F}_{\mathrm{T}}) > (1-t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$ and $tR_{1,m}(\mathbf{F}_{\mathrm{T}}) < (1-t)R_{2,m,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$. When this happens, it is always possible to improve the end-to-end sum-rate performance of the system design. For example, I can always fix all the other transmit precoders and scale down the norm of $\mathbf{F}_{\mathrm{T},k}$ to obtain a new set of transmit precoders $(\mathbf{F}'_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}})$ so that $tR_{1,k}(\mathbf{F}'_{\mathrm{T}}) = (1-t)R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}})$. Note that this decreases the interference power from transmitter k to all other relays on the first hop, improving the achievable rates to all other relays, especially $tR_{1,m}(\mathbf{F}'_{\mathrm{T}}) > tR_{1,m}(\mathbf{F}_{\mathrm{T}})$. This means that $R_{\mathrm{sum}}(\mathbf{F}'_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}}) > R_{\mathrm{sum}}(\mathbf{F}_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}})$. Thus, an efficient transmit precoder design in terms of end-to-end sum-rate maximization should not cause any two-hop rate mismatch.

4.4 Transmit Beamforming Design

In this section, I propose an algorithm for finding high-quality suboptimal solutions to $(\mathcal{OP})$. Subsection 4.4.1 discusses briefly the design of $\mathbf{F}_{\mathrm{X}}$. The design of $\mathbf{F}_{\mathrm{T}}$ is presented in Subsection 4.4.2.

4.4.1 Second-Hop Transmit Beamforming Design

I design $\mathbf{F}_{\mathrm{X}}$ by treating the transmission on the second hop as the single-hop interference broadcast channel. The optimization problem for designing $\mathbf{F}_{\mathrm{X}}$ is formulated as follows

$$
\begin{aligned}
(\mathcal{SP}):\quad & \max_{\mathbf{F}_{\mathrm{X}} \in \mathbb{F}_{\mathrm{X}}} && \sum_{k \in \mathcal{K}_{\mathrm{X}}} R_{2,k,\mathrm{sum}}(\mathbf{F}_{\mathrm{X}}) \\
& \text{subject to} && \sum_{\substack{m \in \mathcal{K}_{\mathrm{R}} \\ \chi(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{X},m}\mathbf{F}_{\mathrm{X},m}^{*}) \le p_{\mathrm{X},k}, \forall k \in \mathcal{K}_{\mathrm{X}}.
\end{aligned} \tag{4.16}
$$

Note that $(\mathcal{SP})$ is nonconvex and NP-hard. Nevertheless, its stationary points can be found by existing algorithms for the single-hop interference broadcast channel. The principle of many existing algorithms is to formulate a series of related optimization problems that are readily solvable by available methods in polynomial time and provide multiple approximations or relaxations of the original sum-utility problem. In general, the globally optimal solutions of these related problems converge to the stationary points of the original sum-utility maximization problem. The key requirement for the applicability of existing algorithms is that the utility function of the original problem is continuously differentiable at every point. An example is the algorithm for transmit precoder design via matrix-weighted sum-MSE minimization in [165, 187]. The details of the algorithm are presented in [165, 187].

Let $\bar{\mathbf{F}}_{\mathrm{X}}$ denote the resulting second-hop transmit precoders. I denote $\bar{\xi}_{2,k} = \xi_{2,k}(\bar{\mathbf{F}}_{\mathrm{X}})$ for $k \in \mathcal{K}_{\mathrm{X}}$. By setting $R_{2,k,\mathrm{sum}}(\bar{\mathbf{F}}_{\mathrm{X}})$ equal to $\log_2(1 + \bar{\xi}_{2,k})$, I obtain

$$\bar{\xi}_{2,k} = 2^{R_{2,k,\mathrm{sum}}(\bar{\mathbf{F}}_{\mathrm{X}})} - 1. \tag{4.17}$$

I assume that each receiver $k \in \mathcal{K}_{\mathrm{R}}$ feeds back the information of $R_{2,k}(\bar{\mathbf{F}}_{\mathrm{X}})$ to its associated relay, i.e., relay $\chi(k)$. Then, I assume that each relay $k \in \mathcal{K}_{\mathrm{X}}$ can compute $R_{2,k,\mathrm{sum}}(\bar{\mathbf{F}}_{\mathrm{X}})$ and $\bar{\xi}_{2,k}$.

4.4.2 First-Hop Transmit Beamforming Design

4.4.2.1 Subproblem Formulation and Challenges

This subsection focuses on designing $\mathbf{F}_{\mathrm{T}}$ given knowledge of t and $\bar{\xi}_{2,k}$ for $k \in \mathcal{K}_{\mathrm{X}}$. It follows from $(\mathcal{OP})$ that the problem for designing $\mathbf{F}_{\mathrm{T}}$ is formulated as follows

$$\begin{aligned}(\mathcal{FP}): \quad \max_{\mathbf{F}_{\mathrm{T}} \in \mathbb{F}_{\mathrm{T}}} \quad & \sum_{k \in \mathcal{K}_{\mathrm{X}}} \min\{t \log_2(1 + \xi_{1,k}(\mathbf{F}_{\mathrm{T}})), (1-t)\log_2(1 + \bar{\xi}_{2,k})\} \\ \text{subject to} \quad & \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m) = k}} \mathrm{tr}(\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^{*}) \le p_{\mathrm{T},k}, \forall k \in \mathcal{K}_{\mathrm{T}}. \end{aligned} \tag{4.18}$$

Note that $(\mathcal{FP})$ belongs to the same class of NP-hard sum-utility maximization problems as $(\mathcal{SP})$ and it is even more complicated than $(\mathcal{SP})$. While the objective function of $(\mathcal{SP})$ depends only on the corresponding transmit precoders, i.e., $\mathbf{F}_{\mathrm{X}}$, that of $(\mathcal{FP})$ depends not only on $\mathbf{F}_{\mathrm{T}}$ but also on t and $\bar{\xi}_k$ for $k \in \mathcal{K}_{\mathrm{X}}$. Moreover, $(\mathcal{FP})$ is a max-min optimization problem.

Remark 4.4.1. It is not possible to apply existing algorithms developed for the single-hop interference broadcast channel to find the stationary points of $(\mathcal{FP})$. As discussed in Section 4.4.1, the applicability of existing algorithms requires that the utility function of sum-utility maximization problems be continuously differentiable at every point. Due to the min operation, however, the utility function of $(\mathcal{FP})$ is not continuously differentiable with respect to $\xi_{1,k}(\mathbf{F}_\mathrm{T})$ at the point that makes $t\log_2(1+\xi_{1,k}(\mathbf{F}_\mathrm{T}))$ equal to $(1-t)\log_2(1+\bar{\xi}_{2,k})$.

Remark 4.4.2. In the naïve approach, the timesharing and second-hop configuration are ignored, leading to an approximation of the objective function $\sum_{k\in\mathcal{K}_\mathrm{X}}\log_2(1+\xi_{1,k}(\mathbf{F}_\mathrm{T}))$. The resulting optimization problem has the same form as $(\mathcal{SP})$. Thus, its stationary points can be found by existing single-hop algorithms.

For notational convenience, I define η_k as the following function of $\bar{\xi}_{2,k}$ and t

$$\eta_k = (1+\bar{\xi}_{2,k})^{\frac{1-t}{t}} - 1. \tag{4.19}$$

Note that $t\log_2(1+\eta_k) = (1-t)\log_2(1+\bar{\xi}_{2,k})$. This means that $\xi_{1,k}(\mathbf{F}_\mathrm{T})$ is equal to η_k when the achievable first-hop rate at relay k matches with the sum of achievable second-hop rates from relay k to its associated receivers. Thus, η_k is the rate-matching received SINR at relay k.

4.4.2.2 Proposed Approach

It is challenging to find the stationary points of a complicated max-min optimization problem like $(\mathcal{FP})$. In the subsection, I aim at finding suboptimal solutions to $(\mathcal{FP})$ with high end-to-end sum-rates. Instead of solving directly $(\mathcal{FP})$, I propose to formulate and solve a new sum-utility maximization problem, which I refer to as $(\mathcal{AFP})$. Having the same constraints as $(\mathcal{FP})$, $(\mathcal{AFP})$ uses an approximation of $\min\{t\log_2(1+\xi_{1,k}(\mathbf{F}_\mathrm{T})),(1-t)\log_2(1+\bar{\xi}_{2,k})\}$ as the utility function. Note that such approximations depend on not only $\mathbf{F}_\mathrm{T}$ but also t and $\bar{\xi}_{2,k}$ for $k\in\mathcal{K}_\mathrm{X}$. Let $u_k(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})$ denote the utility function of $(\mathcal{AFP})$. In addition, I propose to solve $(\mathcal{AFP})$ via iterative minimization of weighted sum-MSEs, the well-established technique that has been used widely in prior work [48, 165, 177, 187].

Some guidelines for selecting $u_k(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})$ are provided. First, it must be twice continuously differentiable with respect to $\xi_{1,k}(\mathbf{F}_\mathrm{T})$ at every point. Second, it must satisfy the following condition

$$(1+\xi_{1,k}(\mathbf{F}_\mathrm{T}))u_k''(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})+2u_k'(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})\geq 0. \tag{4.20}$$

This condition is required to ensure that the resulting iterative algorithm for solving $(\mathcal{AFP})$ is convergent as shown in Proposition 1 in [177]. There are many approximate functions that satisfy the conditions in the guidelines. It is still unclear, however, what is the best approximation.

According to Remark 4.4.1, the utility function of $(\mathcal{FP})$ is not continuously differentiable with respect to $\xi_{1,k}(\mathbf{F}_\mathrm{T})$ at the point $\xi_{1,k}(\mathbf{F}_\mathrm{T})=\eta_k$. I propose an approximation of the utility function of $(\mathcal{FP})$ as follows

$$\begin{aligned}&u_k(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})=\\&\qquad\begin{cases}t\log_2(1+\xi_{1,k}(\mathbf{F}_\mathrm{T})), & \text{if } \xi_{1,k}(\mathbf{F}_\mathrm{T})\leq\eta_k,\\ t\log_2(1+\eta_k)+\frac{t}{\log 2}\left[\exp\left(1-\frac{1+\eta_k}{1+\xi_{1,k}(\mathbf{F}_\mathrm{T})}\right)-1\right], & \text{otherwise.}\end{cases}\end{aligned} \tag{4.21}$$

Note that if $\xi_{1,k}(\mathbf{F}_\mathrm{T})\leq\eta_k$, then $u_k(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})$ is exactly equal to $tR_{2,k}(\mathbf{F}_\mathrm{T})$ and hence equal to $R_k(\mathbf{F}_\mathrm{T},\bar{\mathbf{F}}_\mathrm{X})$. If $\xi_{1,k}(\mathbf{F}_\mathrm{T})>\eta_k$, then $u_k(\xi_{1,k}(\mathbf{F}_\mathrm{T}),t,\bar{\xi}_{2,k})$ is larger than both $tR_{2,k}(\mathbf{F}_{\mathrm{T}})$ and $R_k(\mathbf{F}_{\mathrm{T}}, \bar{\mathbf{F}}_{\mathrm{X}})$. Moreover, when $\xi_{1,k}(\mathbf{F}_{\mathrm{T}}) > \eta_k$, the gap between $u_k(\xi_{1,k}(\mathbf{F}_{\mathrm{T}}), t, \bar{\xi}_{2,k})$ and $R_k(\mathbf{F}_{\mathrm{T}}, \bar{\mathbf{F}}_{\mathrm{X}})$ increases with $\xi_{1,k}(\mathbf{F}_{\mathrm{T}})$ and is upper bounded by $t/\log 2$. Using the approximation $u_k(\xi_{1,k}(\mathbf{F}_{\mathrm{T}}), t, \bar{\xi}_{2,k})$ in (4.21) as the utility function, I formulate $(\mathcal{AFP})$ as follows

$$\begin{aligned}(\mathcal{AFP}): \max_{\mathbf{F}_{\mathrm{T}} \in \mathbb{F}_{\mathrm{T}}} \quad & \sum_{k \in \mathcal{K}_{\mathrm{X}}} u_k(\xi_{1,k}(\mathbf{F}_{\mathrm{T}}), t, \bar{\xi}_{2,k}) \\ \text{subject to} \quad & \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m)=k}} \operatorname{tr}(\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^*) \le p_{\mathrm{T},k}, \forall k \in \mathcal{K}_{\mathrm{T}}. \end{aligned} \tag{4.22}$$

It is expected that the stationary points of $(\mathcal{AFP})$ correspond to high-quality suboptimal solutions to $(\mathcal{FP})$.

4.4.2.3 Sum-Utility Maximization via Matrix-Weighted Sum-MSE Minimization

I develop an algorithm for solving $(\mathcal{AFP})$ via an iterative minimization of weighted sum-MSEs. I denote $\mathbf{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k})$ as the MSE matrix at relay $k \in \mathcal{K}_{\mathrm{X}}$, which is given by

$$\begin{aligned}\mathbf{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k}) &= \operatorname{tr}\left(\mathbf{W}_{\mathrm{X},k}^* \left[\mathbf{H}_{k,\mu(k)}\mathbf{F}_{\mathrm{T},k}\mathbf{F}_{\mathrm{T},k}^*\mathbf{H}_{k,\mu(k)}^* + \mathbf{R}_{\mathrm{X},k}(\mathbf{F}_{\mathrm{T}})\right]\mathbf{W}_{\mathrm{X},k}\right) \\ &\quad - \operatorname{tr}(\mathbf{W}_{\mathrm{X},k}^*\mathbf{H}_{k,\mu(k)}\mathbf{F}_{\mathrm{T},k}) - \operatorname{tr}(\mathbf{F}_{\mathrm{T},k}^*\mathbf{H}_{k,\mu(k)}^*\mathbf{W}_{\mathrm{X},k}) + d_{1,k}.\end{aligned}$$

The MSE of the estimate of $\mathbf{x}_{1,k}$ based on $\mathbf{W}_{\mathrm{X},k}^*\mathbf{y}_{\mathrm{X},k}$ is given by

$$\begin{aligned}\varepsilon_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k}) &= \mathbb{E}(\|\mathbf{x}_{1,k} - \mathbf{W}_{\mathrm{X},k}^*\mathbf{y}_{\mathrm{X},k}\|_{\mathrm{F}}^2) \\ &= \operatorname{tr}(\mathbf{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k})).\end{aligned} \tag{4.23}$$

Fixing $\mathbf{F}_{\mathrm{T}}$ and solving $\frac{\partial \mathbb{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k})}{\partial \mathbf{W}_{\mathrm{X},k}^*} = 0$, I can check that $\varepsilon_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k})$ is minimized by the linear MMSE receive filter $\mathbf{W}_{\mathrm{X},k}^{\mathrm{MMSE}}$ given in (4.4). I denote

$\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}}) = \mathbf{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k}^{\mathrm{MMSE}})$. After some manipulation, I obtain the following well-known relationship between $R_{2,k}(\mathbf{F}_{\mathrm{T}})$ and $\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})$

$$R_{2,k}(\mathbf{F}_{\mathrm{T}}) = -\log_2(\det(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}}))). \tag{4.24}$$

Equivalently, I have

$$\xi_{1,k}(\mathbf{F}_{\mathrm{T}}) = 1/\det(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})) - 1. \tag{4.25}$$

Based on (4.25), I define

$$\begin{aligned} & g_k(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})) \\ \triangleq \quad & -\frac{\log 2}{t} u_k\left(1/\det(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})) - 1, t, \bar{\xi}_{2,k}\right) && (4.26) \\ = \quad & \begin{cases} \log\left(\det(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}}))\right), \text{ if } \det(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})) \geq (1+\eta_k)^{-1}, \\ -\log(1+\eta_k) - \exp\left[1 - (1+\eta_k)\det(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}}))\right] + 1, \text{ otherwise.} \end{cases} && (4.27) \end{aligned}$$

It can be checked that $g_k(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}}))$ is twice continuously differentiable with respect to $\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})$ at any point. Moreover, $g_k(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}}))$ is a strictly concave function of $\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})$. Note that $(\mathcal{AFP})$ is equivalent to the following optimization problem

$$\begin{aligned} (\mathcal{GFP}): \min_{\mathbf{F}_{\mathrm{T}} \in \mathbb{F}_{\mathrm{T}}} \quad & \sum_{k \in \mathcal{K}_{\mathrm{X}}} g_k(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})) \\ \text{subject to} \quad & \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^*) \leq p_{\mathrm{T},k}, \forall k \in \mathcal{K}_{\mathrm{T}}. \end{aligned} \tag{4.28}$$

I define the following function

$$\alpha(x) = \begin{cases} 1, \text{ if } x \geq (1+\eta_k)^{-1} \\ (1+\eta_k)x\exp\left[(1+\eta_k)x - 1\right], \text{ otherwise.} \end{cases} \tag{4.29}$$

The first-order gradient of $g_k(\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T}))$ with respect to $\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T})$ is given by

$$\begin{aligned}\nabla g_k(\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T})) &\triangleq \frac{\partial g_k(\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T}))}{\partial \mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T})}\\ &= \alpha(\det(\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T})))(\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T}))^{-1}. \end{aligned} \tag{4.30}$$

According to Theorem 2 in [187], the inverse mapping of $\nabla g_k(\mathbf{E}_{k,0}(\mathbf{F}_\mathrm{T}))$ is well-defined. I refer to it as $\gamma_k(\cdot) : \mathbb{C}^{d_{1,k}\times d_{1,k}} \to \mathbb{C}^{d_{1,k}\times d_{1,k}}$.

I now use the technique in [165, 187] to solve $(\mathcal{AFP})$ via matrix-weighted sum-MSE minimization. I introduce auxiliary variables $\mathbf{V}_k \in \mathbb{C}^{d_{1,k}\times d_{1,k}}$ for $k \in \mathcal{K}_\mathrm{X}$. I define $\mathbf{V} \triangleq (\mathbf{V}_1, \cdots, \mathbf{V}_{K_\mathrm{X}}) \in \mathbb{V} \triangleq \mathbb{C}^{d_{1,1}\times d_{1,1}} \times \cdots \times \mathbb{C}^{d_{1,K_\mathrm{X}}\times d_{1,K_\mathrm{X}}}$. I define the following matrix-weighted sum-MSE function

$$s_k(\mathbf{F}_\mathrm{T}, \mathbf{W}_\mathrm{X}, \mathbf{V}) = \mathrm{tr}(\mathbf{V}_k^*\mathbf{E}_k(\mathbf{F}_\mathrm{T}, \mathbf{W}_{\mathrm{X},k})) + g_k(\gamma_k(\mathbf{V}_k)) - \mathrm{tr}(\mathbf{V}_k^*)\gamma_k(\mathbf{V}_k). \tag{4.31}$$

A matrix-weighted sum-MSE minimization problem is formulated as follows

$$\begin{aligned}(\mathcal{MFP}) : \quad \min_{(\mathbf{F}_\mathrm{T},\mathbf{W}_\mathrm{X},\mathbf{V})\in\mathbb{F}_\mathrm{T}\times\mathbb{W}_\mathrm{X}\times\mathbb{V}} \quad & \sum_{k\in\mathcal{K}_\mathrm{X}} s_k(\mathbf{F}_\mathrm{T}, \mathbf{W}_\mathrm{X}, \mathbf{V})\\ \text{subject to} \quad & \sum_{\substack{m\in\mathcal{K}_\mathrm{X}\\ \mu(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{T},m}\mathbf{F}_{\mathrm{T},m}^*) \le p_{\mathrm{T},k}, \forall k \in \mathcal{K}_\mathrm{T}.\end{aligned}$$

It follows immediately from Theorem 2 in [187] that $(\mathcal{MFP})$ and $(\mathcal{GFP})$ have the same stationary points if the relays use their corresponding linear MMSE receive filters and the matrix weights $\mathbf{V}_k = \nabla g_k(\mathbf{E}_k(\mathbf{F}_\mathrm{T}, \mathbf{W}_\mathrm{X}))$ for any $\mathbf{F}_\mathrm{T}$ and $\mathbf{W}_\mathrm{X}$. Thus, I can find the stationary points of $(\mathcal{AFP})$ by solving $(\mathcal{GFP})$.

4.4.2.4 Algorithm for Matrix-Weighted Sum-MSE Minimization

I adopt an alternating minimization approach to develop an iterative algorithm for finding the stationary points of $(\mathcal{MFP})$. In each iteration, I focus on determining only one of the sets of parameters $\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X}}$, and $\mathbf{V}$ while assuming the others are fixed. When $\mathbf{F}_{\mathrm{X}}$ and $\mathbf{V}$ are fixed, the optimal linear receive filter at relay $k \in \mathcal{K}_{\mathrm{X}}$ is exactly $\mathbf{W}_{\mathrm{X},k}^{\mathrm{MMSE}}$ given in (4.4). In addition, as discussed in Subsection 4.4.2.3, when $\mathbf{F}_{\mathrm{T}}$ and $\mathbf{W}_{\mathrm{X}}$ are fixed, the optimal matrix weights are as follows

$$\mathbf{V}_k^{\mathrm{opt}} = \nabla g_k(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})), \forall k \in \mathcal{K}_{\mathrm{X}}. \tag{4.32}$$

It follows from (4.23) that $\mathbf{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k})$ is a Hermitian and positive semidefinite matrix for any $\mathbf{F}_{\mathrm{T}}$ and $\mathbf{W}_{\mathrm{X}}$. Combined with (4.30), I have $\nabla g_k(\mathbf{E}_{k,0}(\mathbf{F}_{\mathrm{T}})$ is a Hermitian and positive semidefinite matrix. This means that if I always choose $\mathbf{V}_k = \mathbf{V}_k^{\mathrm{opt}}$ according to (4.32), then $\mathbf{V}_k$ is a Hermitian and positive semidefinite matrix for $k \in \mathcal{K}_{\mathrm{X}}$.

What remains is the design of $\mathbf{F}_{\mathrm{T}}$ when $\mathbf{W}_{\mathrm{X}}$ and $\mathbf{V}_{\mathrm{X}}$ are fixed. When $\mathbf{W}_{\mathrm{X}}$ and $\mathbf{V}_{\mathrm{X}}$ are fixed, I need to solve the following optimization problem to determine $\mathbf{F}_{\mathrm{T}}$

$$\begin{aligned} (\mathcal{MFP}\text{-}\mathcal{F}_{\mathrm{T}}): \quad \min_{\mathbf{F}_{\mathrm{T}} \in \mathbb{F}_{\mathrm{T}}} \quad & \sum_{k \in \mathcal{K}_{\mathrm{X}}} \mathrm{tr}(\mathbf{V}_k^* \mathbf{E}_k(\mathbf{F}_{\mathrm{T}}, \mathbf{W}_{\mathrm{X},k})) \\ \text{subject to} \quad & \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m)=k}} \mathrm{tr}(\mathbf{F}_{\mathrm{T},m} \mathbf{F}_{\mathrm{T},m}^*) \leq p_{\mathrm{T},k}, \forall k \in \mathcal{K}_{\mathrm{T}}. \end{aligned} \tag{4.33}$$

Note that $(\mathcal{MFP}\text{-}\mathcal{F}_{\mathrm{T}})$ is convex with respect to $\mathbf{F}_{\mathrm{T},k}$ for $k \in \mathcal{K}_{\mathrm{X}}$. Let $\lambda_k \geq 0$ be the Lagrangian multiplier associated with the sum transmit power constraint at transmitter $k \in \mathcal{K}_{\mathrm{T}}$. Based on the optimality condition of $(\mathcal{MFP}\text{-}\mathcal{F}_{\mathrm{T}})$, the globally optimal solution of ($\mathcal{MFP}$-$\mathcal{F}_{\mathrm{T}}$) must has the following form for $k \in \mathcal{K}_{\mathrm{X}}$

$$\begin{aligned} \mathbf{F}_{\mathrm{T},k}(\lambda_k) &= \left(\sum_{m \in \mathcal{K}_{\mathrm{X}}} \mathbf{H}^*_{m,\mu(k)} \mathbf{W}_{\mathrm{X},m} \mathbf{V}_m \mathbf{W}^*_{\mathrm{X},m} \mathbf{H}^*_{m,\mu(k)} + \lambda_k \mathbf{I}_{N_{\mathrm{T},k}} \right)^{-1} \\ &\quad \times \mathbf{H}^*_{k,\mu(k)} \mathbf{W}_{\mathrm{X},k} \mathbf{V}_k. \end{aligned} \tag{4.34}$$

Since $\mathbf{V}_k$ is a Hermitian and positive semidefinite matrix, the following matrix $\sum_{m \in \mathcal{K}_{\mathrm{X}}} \mathbf{H}^*_{m,\mu(k)} \mathbf{W}_{\mathrm{X},m} \mathbf{V}_m \mathbf{W}^*_{\mathrm{X},m} \mathbf{H}^*_{m,\mu(k)}$ is also Hermitian and positive semidefinite. It can be checked that $\mathrm{tr}\left(\mathbf{F}_{\mathrm{T},k}(\lambda_k)\mathbf{F}^*_{\mathrm{T},k}(\lambda_k)\right)$ is strictly decreasing with λ_k in $[0, +\infty)$. The optimal Lagrangian multiplier $\lambda_k^* \geq 0$ is chosen such that the complementary slackness condition of the sum power constraint at transmitter $k \in \mathcal{K}_{\mathrm{T}}$ is satisfied. For any $k \in \mathcal{K}_{\mathrm{T}}$, if $\sum_{m \in \mathcal{K}_{\mathrm{X}}:\ \mu(m)=k} \mathrm{tr}(\mathbf{F}_{\mathrm{T},k}(0)\mathbf{F}^*_{\mathrm{T},k}(0)) \leq p_{\mathrm{T},k}$, then $\mathbf{F}^{\mathrm{opt}}_{\mathrm{T},k} = \mathbf{F}_{\mathrm{T},k}(0)$. Otherwise, λ_k^* is the unique solution of the following equation

$$\sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ \mu(m)=k}} \mathrm{tr}\left(\mathbf{F}_{\mathrm{T},k}(\lambda_k)\mathbf{F}^*_{\mathrm{T},k}(\lambda_k)\right) = p_{\mathrm{T},k}. \tag{4.35}$$

This equation can be solved by using one-dimensional search techniques, e.g., the bisection method.

Note that in the proposed alternating minimization algorithm for solving ($\mathcal{MFP}$), I am able to find the globally optimal solutions to the corresponding optimization problem in each iteration. Therefore, the algorithm is guaranteed to converge to a stationary point of ($\mathcal{MFP}$), which is also a stationary point of ($\mathcal{AFP}$). Let $\bar{\mathbf{F}}_{\mathrm{T}}$ denote the transmit precoders corresponding to the resulting stationary point. It is worth to emphasize that it is not guaranteed that I can find a stationary point of ($\mathcal{FP}$). Thus, a two-hop rate mismatch may still happen for the resulting suboptimal solution $(\bar{\mathbf{F}}_{\mathrm{T}}, \bar{\mathbf{F}}_{\mathrm{X}})$ of the original transmit precoder design problem ($\mathcal{OP}$). This leaves room for potential improvements in terms of maximizing $R_{\mathrm{sum}}(\mathbf{F}_{\mathrm{T}}, \bar{\mathbf{F}}_{\mathrm{X}})$.

4.4.2.5 Rate-Matching Transmit Power Control

I propose an iterative power control method for eliminating any residual two-hop rate mismatch corresponding to $(\bar{\mathbf{F}}_{\mathrm{T}}, \bar{\mathbf{F}}_{\mathrm{X}})$. Let $\mathbf{F}_{\mathrm{T},k}^{(n)}$ denote the transmit precoder for the transmission to relay $k \in \mathcal{K}_{\mathrm{X}}$ in iteration n of this method. Note that $\mathbf{F}_{\mathrm{T},k}^{(0)} = \bar{\mathbf{F}}_{\mathrm{T},k}$ for $k \in \mathcal{K}_{\mathrm{X}}$. Let $\theta_k^{(n)} \in \mathbb{R}$ be the norm of $\mathbf{F}_{\mathrm{T},k}^{(n)}$, i.e., the power allocated for the transmission from transmitter $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ to relay $k \in \mathcal{K}_{\mathrm{X}}$ in iteration n. I propose to fix the shapes of the transmit precoders and to adjust only their norm θ_k for $k \in \mathcal{K}_{\mathrm{X}}$. It follows that $\mathbf{F}_{\mathrm{T},k}^{(n)} = \theta_k^{(n)} \frac{\bar{\mathbf{F}}_{\mathrm{T},k}}{\|\bar{\mathbf{F}}_{\mathrm{T},k}\|_{\mathrm{F}}}$ for all $n \geq 0$ where $\frac{\bar{\mathbf{F}}_{\mathrm{T},k}}{\|\bar{\mathbf{F}}_{\mathrm{T},k}\|_{\mathrm{F}}}$ has unit norm and represents the shape of $\bar{\mathbf{F}}_{\mathrm{T},k}$.

I define the following notation $\boldsymbol{\theta}^{(n)} = (\theta_1^{(n)}, \cdots, \theta_{K_{\mathrm{X}}}^{(n)}) \in \mathbb{R}^{K_{\mathrm{X}} \times 1}$ and $\boldsymbol{\theta}_{-k}^{(n)} = (\theta_1^{(n)}, \cdots, \theta_{k-1}^{(n)}, \theta_{k+1}^{(n)} \cdots, \theta_{K_{\mathrm{X}}}^{(n)})$. $\boldsymbol{\theta}^{(n)}$ and $(\theta_k^{(n)}; \boldsymbol{\theta}_{-k}^{(n)})$ are used interchangeably in the subsection. Also, I denote $\boldsymbol{\mathcal{H}}_{m,\mu(k)} = \mathbf{H}_{m,\mu(k)} \frac{\bar{\mathbf{F}}_{\mathrm{T},k}}{\|\bar{\mathbf{F}}_{\mathrm{T},k}\|_{\mathrm{F}}}$ for $k, m \in \mathcal{K}_{\mathrm{X}}$. The interference plus noise covariance matrix at relay $k \in \mathcal{K}_{\mathrm{X}}$ is rewritten as follows

$$\mathbf{R}_{\mathrm{X},k}(\boldsymbol{\theta}_{-k}^{(n)}) = \sum_{\substack{m \in \mathcal{K}_{\mathrm{X}} \\ m \neq k}} \theta_m^{(n)} \boldsymbol{\mathcal{H}}_{k,\mu(m)} \boldsymbol{\mathcal{H}}_{k,\mu(m)}^{*} + \sigma_{\mathrm{X},k}^2 \mathbf{I}_{N_{\mathrm{X},k}}. \tag{4.36}$$

The maximum achievable rate at relay $k \in \mathcal{K}_{\mathrm{X}}$ is rewritten as follows

$$R_{1,k}(\boldsymbol{\theta}^{(n)}) \;=\; \log_2 \det \Big(\mathbf{I} + \theta_k^{(n)} \boldsymbol{\mathcal{H}}_{k,\mu(k)}^{*} \big[\mathbf{R}_{\mathrm{X},k}(\boldsymbol{\theta}_{-k}^{(n)}) \big]^{-1} \boldsymbol{\mathcal{H}}_{k,\mu(k)} \Big). \tag{4.37}$$

Thus, it follows that

$$\xi_{1,k}(\boldsymbol{\theta}^{(n)}) = \det\left(\mathbf{I} + \theta_k^{(n)} \boldsymbol{\mathcal{H}}_{k,\mu(k)}^* \left[\mathbf{R}_{\mathrm{X},k}(\boldsymbol{\theta}_{-k}^{(n)})\right]^{-1} \boldsymbol{\mathcal{H}}_{k,\mu(k)}\right) - 1. \tag{4.38}$$

The end-to-end achievable rate corresponding to relay $k \in \mathcal{K}_{\mathrm{X}}$ is written as

$$R_k(\boldsymbol{\theta}^{(n)}) = \min\{tR_{1,k}(\boldsymbol{\theta}^{(n)}), (1-t)\log_2(1+\bar{\xi}_{2,k})\}. \tag{4.39}$$

Note that $\boldsymbol{\mathcal{H}}_{k,\mu k}^*[\mathbf{R}_{\mathrm{X},k}(\boldsymbol{\theta}_{-k}^{(n)})]^{-1}\boldsymbol{\mathcal{H}}_{k,\mu k}$ is independent of $\theta_k^{(n)}$. It is also a Hermitian and positive semidefinite matrix. Since $\det(\mathbf{I} + x\mathbf{A})$ is strictly increasing in x for $x \geq 0$ when $\mathbf{A}$ is a positive semidefinite matrix, both $R_{1,k}(\boldsymbol{\theta}^{(n)})$ and $\xi_{1,k}(\boldsymbol{\theta}^{(n)})$ are strictly increasing in $\theta_k^{(n)}$ when $\theta_k^{(n)} \geq 0$.

Let $\mathcal{A}^{(n)} \triangleq \{k \in \mathcal{K}_{\mathrm{X}} : \ tR_{1,k}(\boldsymbol{\theta}^{(n)}) > (1-t)\log_2(1+\bar{\xi}_{2,k})\}$ be the index set of the relays with a dominant first hop in iteration n. If $k \in \mathcal{A}^{(n)}$, then excess power is allocated for the transmission to relay k. Similarly, let $\mathcal{B}^{(n)} \triangleq \{k \in \mathcal{K}_{\mathrm{X}} : \ tR_{1,k}(\boldsymbol{\theta}^{(n)}) < (1-t)\log_2(1+\bar{\xi}_{2,k})\}$ be the index set of the relays with a dominant second hop in iteration n. A two-hop rate mismatch happens if and only if $\mathcal{A}^{(n)} \not\equiv \emptyset$ and $\mathcal{B}^{(n)} \not\equiv \emptyset$. When a two-hop rate mismatch happens, I consider an arbitrary $k_{\mathcal{A}} \in \mathcal{A}^{(n)}$ and $k_{\mathcal{B}} \in \mathcal{B}^{(n)}$. It follows $\mathcal{A}^{(n)} \cap \mathcal{B}^{(n)} \equiv \emptyset$ that $k_{\mathcal{A}} \neq k_{\mathcal{B}}$. When $\theta_m^{(n)}$ is fixed for $m \in \mathcal{K}_{\mathrm{X}}$ and $m \neq k_{\mathcal{A}}$, since $k_{\mathcal{A}} \in \mathcal{A}^{(n)}$, there must exist $\phi_{k_{\mathcal{A}}}^{(n)} \in (0, \theta_{k_{\mathcal{A}}}^{(n)})$ such that

$$tR_{1,k}(\phi_{k_{\mathcal{A}}}^{(n)}; \boldsymbol{\theta}_{-k_{\mathcal{A}}}^{(n)})) = (1-t)\log_2(1+\bar{\xi}_{2,k_{\mathcal{A}}}). \tag{4.40}$$

Equivalently, it follows that

$$\det\left(\mathbf{I} + \phi_{k_{\mathcal{A}}}^{(n)} \boldsymbol{\mathcal{H}}_{k_{\mathcal{A}},\mu(k_{\mathcal{A}})}^* \left[\mathbf{R}_{\mathrm{X},k_{\mathcal{A}}}(\boldsymbol{\theta}_{-k_{\mathcal{A}}}^{(n)})\right]^{-1} \boldsymbol{\mathcal{H}}_{k_{\mathcal{A}},\mu(k_{\mathcal{A}})}\right) - 1 = \eta_{k_{\mathcal{A}}}. \tag{4.41}$$

Since the left-hand side of this equation is strictly increasing in $\phi_{k_{\mathcal{A}}}^{(n)}$, this equation has a unique solution, which can be found by using one dimensional search techniques, e.g., the bisection method. Note that $\phi_{k_{\mathcal{A}}}^{(n)}$ is the rate-matching power for the transmission to relay $k_{\mathcal{A}}$ in iteration n.

The key observation for the proposed power control algorithm is that if a two-hop rate mismatch happens at the end of iteration n, excess power can be reduced by setting $\theta_{k_{\mathcal{A}}}^{(n+1)} = \phi_{k_{\mathcal{A}}}$ and $\boldsymbol{\theta}_{-k_{\mathcal{A}}}^{(n+1)} = \boldsymbol{\theta}_{-k_{\mathcal{A}}}^{(n)}$ for an arbitrary $k_{\mathcal{A}} \in \mathcal{A}^{(n)}$. Note that $R_{k_{\mathcal{A}}}(\boldsymbol{\theta}^{(n+1)}) = (1-t)\log_2(1+\bar{\xi}_{2,k_{\mathcal{A}}}) = R_{k_{\mathcal{A}}}(\boldsymbol{\theta}^{(n)})$, Also, this excess power reduction decreases the power of interference observed by all relay $m \neq k_{\mathrm{A}}$, leading to $R_{1,m}(\boldsymbol{\theta}^{(n+1)}) \geq R_{1,m}(\boldsymbol{\theta}^{(n)})$ and hence $R_m(\boldsymbol{\theta}^{(n+1)}) \geq R_m(\boldsymbol{\theta}^{(n)})$. Especially, it follows that $R_{1,k_{\mathcal{B}}}(\boldsymbol{\theta}^{(n+1)}) > R_{1,k_{\mathcal{B}}}(\boldsymbol{\theta}^{(n)})$ and hence $R_{k_{\mathcal{B}}}(\boldsymbol{\theta}^{(n+1)}) > R_{k_{\mathcal{B}}}(\boldsymbol{\theta}^{(n)})$. As a result, the end-to-end sum-rate is strictly improved, i.e., $\sum_{k\in\mathcal{K}_{\mathrm{X}}} R_k(\boldsymbol{\theta}^{(n)}) < \sum_{k\in\mathcal{K}_{\mathrm{X}}} R_k(\boldsymbol{\theta}^{(n+1)})$. Thus, when a two-hop rate mismatch happens, reducing excess power in a controlled manner strictly increases the sum of achievable end-to-end rates.

Based on the observation, I propose an iterative algorithm for updating the power allocated for the transmission from the transmitters to the relays. Specifically, at the end of each iteration $n \geq 0$, each relay $k \in \mathcal{K}_{\mathrm{R}}$ computes $\xi_{1,k}(\boldsymbol{\theta}^{(n)})$ to check if the transmission to itself is allocated excess power. If $\xi_{1,k}(\boldsymbol{\theta}^{(n)}) \leq \eta_k$, then the power allocated for the transmission to relay k does not need to change, i.e., $\theta_k^{(n+1)} = \theta_k^{(n)}$. Otherwise, relay k determines the corresponding rate-matching power $\phi_k^{(n)}$ and feeds back the value to its associated transmitter $\mu(k)$ to instruct the transmitter to update

$\theta_k^{(n+1)} = \phi_k^{(n+1)} \in (0, \theta_k^{(n)})$. In other words, the power update rule for $k \in \mathcal{K}_{\mathrm{X}}$ and $n \geq 0$ is as follows

$$\theta_k^{(n+1)} = \begin{cases} \phi_k^{(n)}, & \text{if } k \in \mathcal{A}^{(n)}, \\ \theta_k^{(n)}, & \text{otherwise.} \end{cases} \tag{4.42}$$

This process is repeated until the algorithm converges or the maximum number of iteration is reached. It follows from the power update rule that $\theta_k^{(n+1)} \leq \theta_k^{(n)}$ for $k \in \mathcal{K}_{\mathrm{X}}$ and $n \geq 0$. Several properties of the proposed algorithm are stated and proved in Proposition 4.4.1, Proposition 4.4.2, and Proposition 4.4.3.

Proposition 4.4.1. *The proposed power control algorithm is guaranteed to converge.*

Proof. Note that $\theta_k^{(n+1)} \leq \theta_k^{(n)}$ for $k \in \mathcal{K}_{\mathrm{X}}$ and $n \geq 1$. Since $\theta_k^{(n)}$ is nonnegative, it is lower bounded by 0. Thus, it is guaranteed that the proposed power control algorithm converge as the number of iterations goes to infinity. □

Proposition 4.4.2. *The resulting solution of the proposed power control algorithm does not causes a two-hop rate mismatch.*

Proof. Let n_0 be the index of the iteration at which the proposed algorithm is convergent, i.e., $\theta_k^{(n_0)} = \theta_k^{(n_0+1)}$ for $k \in \mathcal{K}_{\mathrm{X}}$. This means that $\xi_{1,k}(\boldsymbol{\theta}^{(n_0)}) \leq \eta_k$ for all $k \in \mathcal{K}_{\mathrm{X}}$ or $\mathcal{A}^{(n_0)} \equiv \emptyset$. Recall that a two-hop rate mismatch happens in iteration n if and only if $\mathcal{A}^{(n)} \not\equiv \emptyset$ and $\mathcal{B}^{(n)} \not\equiv \emptyset$. Thus there is not any two-hop rate mismatch in iteration n_0. □

Proposition 4.4.3. *The proposed algorithm does not decrease $R_k(\boldsymbol{\theta}^{(n)})$ over iterations for $n \geq 0$.*

Proof. It can be showed that if $\mathbf{A}$ is positive semidefinite, then $\det(\mathbf{I}+\mathbf{B}^*(\mathbf{I}+x\mathbf{A})^{-1}\mathbf{B})$ is strictly decreasing in x for $x \geq 0$ for any $\mathbf{B}$. From (4.36) and (4.38), I have

$$\begin{aligned}
&\xi_{1,k}(\boldsymbol{\theta}^{(n+1)}) + 1 \\
= \;& \det\left(\mathbf{I} + \theta_k^{(n+1)}\mathcal{H}^*_{k,\mu(k)}\left[\sum_{\substack{m\in\mathcal{K}_\mathrm{X}\\ m\neq k}} \theta_m^{(n+1)}\mathcal{H}_{k,\mu(m)}\mathcal{H}^*_{k,\mu(m)} + \sigma_k^2\mathbf{I}\right]^{-1}\mathcal{H}_{k,\mu(k)}\right) \\
\geq \;& \det\left(\mathbf{I} + \theta_k^{(n+1)}\mathcal{H}^*_{k,\mu(k)}\left[\sum_{\substack{m\in\mathcal{K}_\mathrm{X}\\ m\neq k}} \theta_m^{(n)}\mathcal{H}_{k,\mu(m)}\mathcal{H}^*_{k,\mu(m)} + \sigma_k^2\mathbf{I}\right]^{-1}\mathcal{H}_{k,\mu(k)}\right)
\end{aligned} \tag{4.43}$$

$$= \begin{cases} \eta_k + 1, \text{ if } k \in \mathcal{A}^{(n)} \\ \xi_{1,k}(\boldsymbol{\theta}^{(n)}) + 1, \text{ otherwise,} \end{cases} \tag{4.44}$$

where (4.43) comes from the property that $\theta_k^{(n+1)} \leq \theta_k^{(n)}$ for $k \in \mathcal{K}_\mathrm{X}$ and (4.44) comes from (4.42) and (4.41). It follows that $\min\{\eta_k, \xi_{1,k}^{(n+1)}\} \geq \min\{\eta_k, \xi_{1,k}^{(n)}\}$; equivalently, $R_k(\boldsymbol{\theta}^{(n+1)}) \geq R_k(\boldsymbol{\theta}^{(n)})$ for $k \in \mathcal{K}_\mathrm{X}$. □

Note that the proposed power control algorithm does not help find any stationary points of $(\mathcal{OP})$. Essentially, it helps eliminate the residual two-hop rate mismatch in $(\bar{\mathbf{F}}_\mathrm{T}, \bar{\mathbf{F}}_\mathrm{X})$ to find a potentially better suboptimal solution to $(\mathcal{OP})$.

4.4.3 Distributed Implementation

The proposed transmit precoding algorithm for the relay interference broadcast channel allows for distributed implementation. Similar to [165, 184, 186, 187], two assumptions are needed. First, each transmitting node has the

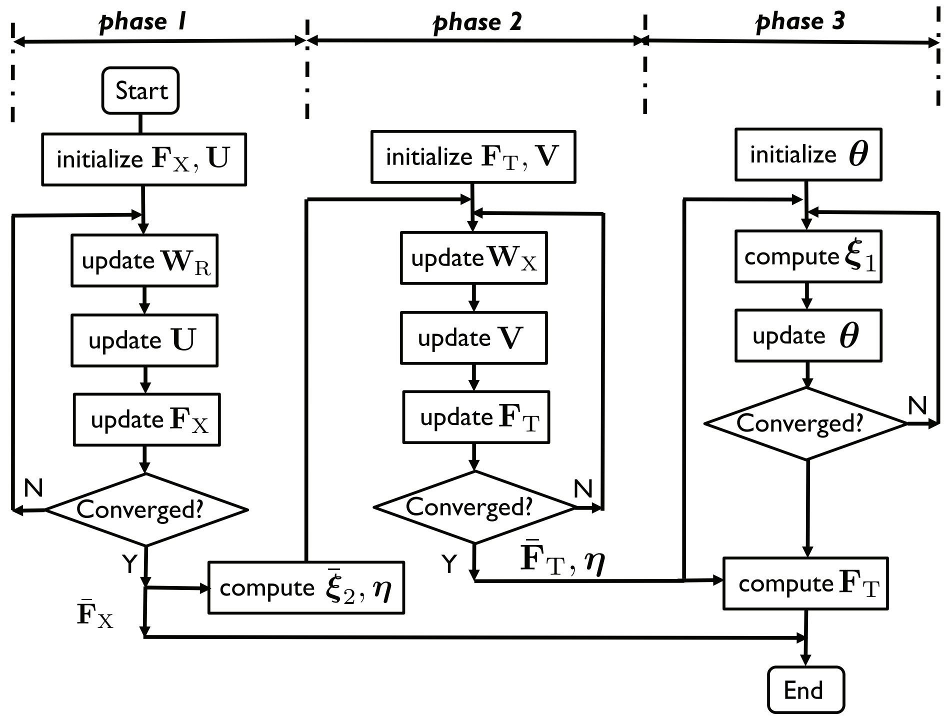

Figure 4.2: Flow diagram of the proposed algorithm for the relay interference broadcast channel.

corresponding local channel state information (CSI). Specifically, on the first hop, transmitter $k \in \mathcal{K}_{\mathrm{T}}$ has the CSI of $\mathbf{H}_{m,k}$ for all $m \in \mathcal{K}_{\mathrm{X}}$; on the second hop, relay $m \in \mathcal{K}_{\mathrm{X}}$ has the CSI of $\mathbf{G}_{q,m}$ for all $q \in \mathcal{K}_{\mathrm{R}}$. Second, there is a feedback channel to send information from a receiving node to its serving node, i.e., from receiver $q \in \mathcal{K}_{\mathrm{R}}$ to relay $\xi(q) \in \mathcal{K}_{\mathrm{X}}$ and from relay $m \in \mathcal{K}_{\mathrm{X}}$ to transmitter $\mu(m) \in \mathcal{K}_{\mathrm{T}}$. The flow diagram of the proposed algorithm is presented in Fig. 4.2, where the notation of main parameters is provided in Table 4.1. Note that Phase 1 is a counterpart of Table I in [187].

Recall that it is not guaranteed that the proposed algorithm can find

Table 4.1: Notation of the main parameters

Notation	Parameters
$\chi(k)$	index of the relay aiding receiver $k \in \mathcal{K}_{\mathrm{R}}$
$\mu(k)$	index of the transmitter aided by relay $k \in \mathcal{K}_{\mathrm{X}}$
$\mathbf{F}_{\mathrm{X},q}$	precoder at relay $\chi(q) \in \mathcal{K}_{\mathrm{X}}$ for transmission to receiver $q \in \mathcal{K}_{\mathrm{R}}$
$\bar{\mathbf{F}}_{\mathrm{X},q}$	resulting transmit precoder at relay $\chi(q) \in \mathcal{K}_{\mathrm{X}}$ for transmission to receiver $q \in \mathcal{K}_{\mathrm{R}}$
$\mathbf{W}_{\mathrm{R},q}$	receive filter at receiver $q \in \mathcal{K}_{\mathrm{R}}$
$\mathbf{U}_q$	matrix weight for MSE at receiver $q \in \mathcal{K}_{\mathrm{R}}$ in the second-hop weighted sum-MSEs
$\bar{\xi}_{2,k}$	effective SINR corresponding to second-hop sum-rates from relay $k \in \mathcal{K}_{\mathrm{X}}$
η_k	rate-matching SINR at relay $k \in \mathcal{K}_{\mathrm{X}}$ on the first hop
$\mathbf{F}_{\mathrm{T},k}$	precoder at transmitter $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ for transmission to relay $k \in \mathcal{K}_{\mathrm{X}}$
$\bar{\mathbf{F}}_{\mathrm{T},k}$	resulting precoder at transmitter $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ for transmission to relay $k \in \mathcal{K}_{\mathrm{X}}$ in the second phase
$\mathbf{W}_{\mathrm{X},k}$	receive filter at relay $k \in \mathcal{K}_{\mathrm{X}}$
$\mathbf{V}_k$	matrix weight for MSE at relay $k \in \mathcal{K}_{\mathrm{X}}$ in the first-hop weighted sum-MSEs
θ_k	Frobenius norm of $\mathbf{F}_{\mathrm{T},k}$ in the third phase, i.e., the power for transmission from transmitter $\mu(k) \in \mathcal{K}_{\mathrm{T}}$ to relay $k \in \mathcal{K}_{\mathrm{X}}$
$\xi_{1,k}$	effective received SINR at relay $k \in \mathcal{K}_{\mathrm{X}}$

optimal solutions to $(\mathcal{OP})$. In fact, the end-to-end sum-rate performance of the resulting solution $(\mathbf{F}_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}})$ depends on the initial solutions in the first two phases. One method for improving the end-to-end sum-rate performance of the algorithm is to use multiple random transmit precoders as initial solutions in the first two phases and then to select the best one in terms of end-to-end sum-rate maximization. Such an opportunistic approach, however, may require more coordination among the nodes and result in more overhead in the network.

4.5 Simulations

This section presents Monte Carlo simulations to investigate the end-to-end sum-rate performance of the proposed algorithm. I consider only symmetric relay systems with $N_{\mathrm{T},k} = N_{\mathrm{T}}$ and $p_{\mathrm{T},k} = p_{\mathrm{T}}$ for $k \in \mathcal{K}_{\mathrm{T}}$; $N_{\mathrm{X},m} = N_{\mathrm{X}}$, $d_{1,m} = d_1$, and $p_{\mathrm{X},m} = p_{\mathrm{X}}$ for $m \in \mathcal{K}_{\mathrm{X}}$; and $N_{\mathrm{R},q} = N_{\mathrm{R}}$ and $d_{2,q} = d_2$ for $q \in \mathcal{K}_{\mathrm{R}}$. Except when it is stated explicitly, I use equal timesharing, i.e., $t = 0.5$, and assume $p_{\mathrm{T}} = p_{\mathrm{X}}$. Also, I consider the case in which each transmitter is aided by the same number of relays, i.e., $K_{\mathrm{X}}/K_{\mathrm{T}}$. Each relay forwards data from its associated transmitter to the same number of receivers, i.e., $K_{\mathrm{R}}/K_{\mathrm{X}}$. I denote the system as $(N_{\mathrm{T}}^{K_{\mathrm{T}}} \times N_{\mathrm{X}}^{K_{\mathrm{X}}} \times N_{\mathrm{R}}^{K_{\mathrm{R}}}, d_1 \times d_2)$. The power values are normalized such that $\sigma_{\mathrm{X},m} = 1$ for $m \in \mathcal{K}_{\mathrm{X}}$ and $\sigma_{\mathrm{R},q} = 1$ for $q \in \mathcal{K}_{\mathrm{R}}$. The channels are flat both in time and in frequency. The channel coefficients on two hops are generated as i.i.d. zero-mean unit-variance complex Gaussian random variables. Path loss is not considered in the simulations, thus the average power values of all cross-links on the same hop are equal to each other.

Remark 4.5.1. Similar to the algorithms proposed in Chapter 3, the proposed algorithm works for any fixed topology of base stations and relays in cellular networks. As discussed in 2.6.1, the actual deployment of cellular networks leads to more random locations of base stations and relays [220]. Future work may use the more tractable model for the topologies of base station locations and relay locations is proposed [13, 65] for investigating the impact of interference, including uncoordinated interference, on my proposed algorithm.

The plots are produced by averaging over 1000 random channel realizations. In each channel realization, the initial transceivers are chosen randomly. The maximum number of iterations in the first two phases is 2000 while that of the last phase is 30. I use the naïve approach as the baseline for comparison. The proposed algorithm and the baseline have the same output at the end of the first phase and use the same initial solution in the second phase. In addition, I adopt an opportunistic approach to improve the end-to-end sum-rate performance. Let N be the number of random initializations in the opportunistic approach. The proposed algorithm is repeated N times with the random initializations and choose the one with the highest achievable end-to-end sum-rates.

4.5.1 First-Hop Transmit Beamforming Design based on Approximate End-to-End Rates

This subsection investigates the benefits of the second phase of the proposed algorithm. Recall that the proposed first-hop transmit beamforming design on the second phase is guaranteed to converge to a stationary point of the optimization problem of maximizing an approximation of achievable end-to-end sum-rates. Fig.4.3 provides the simulation results of the average end-to-end sum-rates achieved after the second phase, i.e., before rate-matching power control, in relative comparison with the baseline for the following three systems: $(2^3 \times 2^6 \times 1^{12}, 1 \times 1)$, $(2^4 \times 2^4 \times 2^4, 1 \times 1)$, and $(4^4 \times 4^8 \times 2^{16}, 2 \times 2)$. I

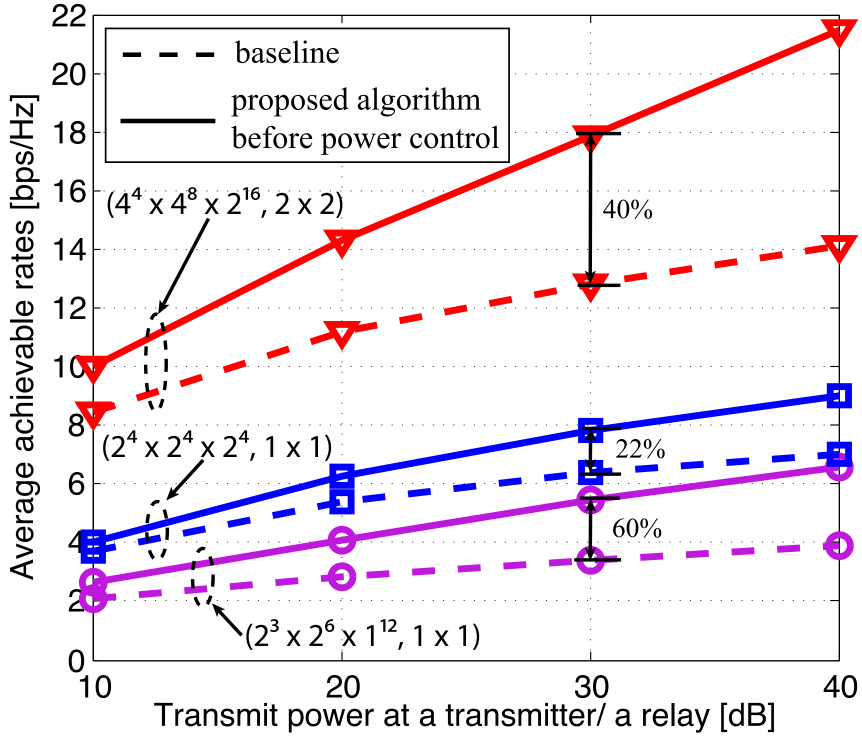

Figure 4.3: Average end-to-end sum-rates of the proposed algorithm after the second phase, i.e., before rate-matching power control, in relative comparison with the baseline.

observe that even before rate-matching power control, the proposed algorithm always outperforms the baseline. At $p_{\mathrm{T}} = p_{\mathrm{X}} = 30$ dB, the gain is 60% for the system $(2^3 \times 2^6 \times 1^{12}, 1 \times 1)$, 22% for the system $(2^4 \times 2^4 \times 2^4, 1 \times 1)$ and 40% for the system $(4^4 \times 4^8 \times 2^{16}, 2 \times 2)$. This means that thanks to the consideration of t and $\xi_{2,k}$ for $k \in \mathcal{K}_{\mathrm{X}}$ in the design of $\mathbf{F}_{\mathrm{T}}$, the second phase of the proposed algorithm is able to alleviate a two-hop rate mismatch to obtain higher end-to-end sum-rates.

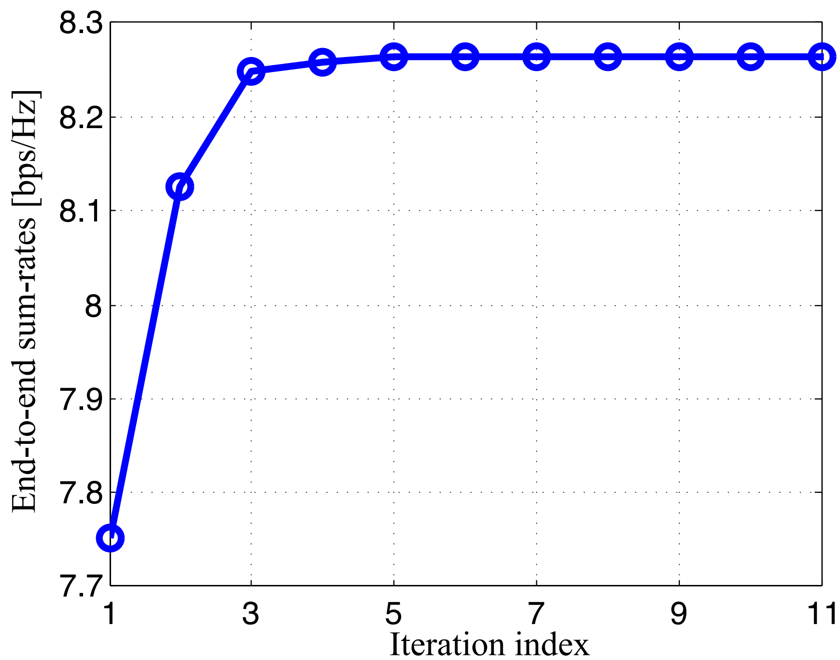

Figure 4.4: Convergence behavior of the rate-matching power control algorithm for $(2^3 \times 2^6 \times 2^{12}, 1 \times 1)$.

4.5.2 Rate-Matching Transmit Power Control

Fig. 4.4 shows the convergence behavior of the proposed rate-matching transmit power control method for a channel realization of the system $(2^3 \times 2^6 \times 2^{12}, 1 \times 1)$. I observe that this method converges in few iterations and it does so monotonically. In addition, as discussed in Subsection 4.4.2.5, the proposed rate-matching power control method can be applied to any $(\mathbf{F}_{\mathrm{T}}, \mathbf{F}_{\mathrm{X}})$.

Fig. 4.5 presents the average achievable end-to-end sum-rates before and after applying the power control method for various strategies. At $p_{\mathrm{T}} = p_{\mathrm{X}} = 30$ dB, the power control method provides a gain of 28% over the output of the baseline. It also provides a gain of 10% over the output of the second

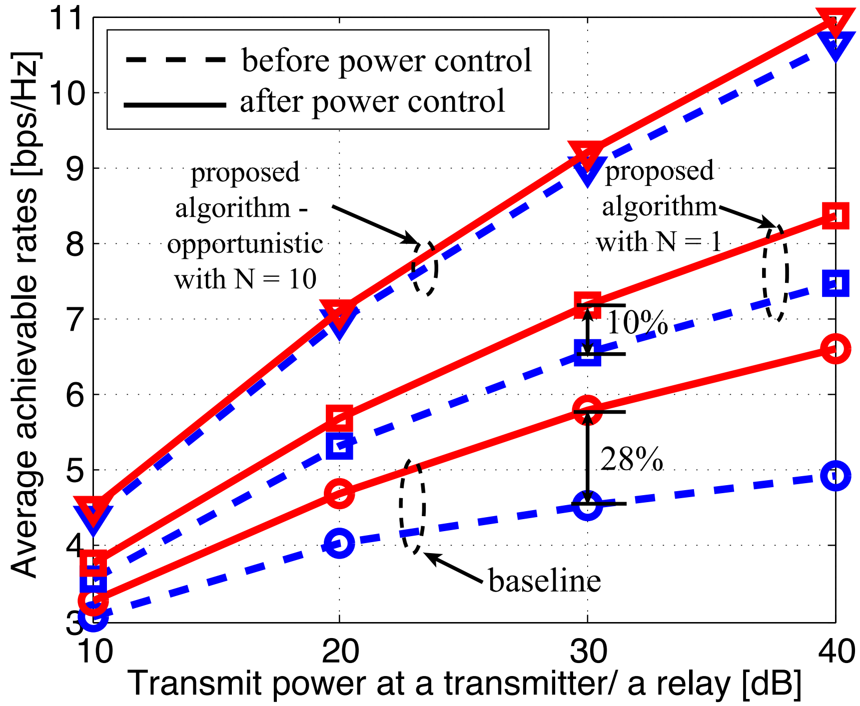

Figure 4.5: Investigating the benefits of the third phase of the proposed algorithm, i.e., rate-matching power control, for $(2^3 \times 2^6 \times 2^{12}, 1 \times 1)$. N is the number of random initializations used in the opportunistic approach.

phase of the proposed three-phase algorithm. If I consider the opportunistic solution of the second phase with $N = 10$, however, the gain provided by the power control method is negligible. Thus, as expected, this power control method is more beneficial when there is a large two-hop rate mismatch in its initial solution.

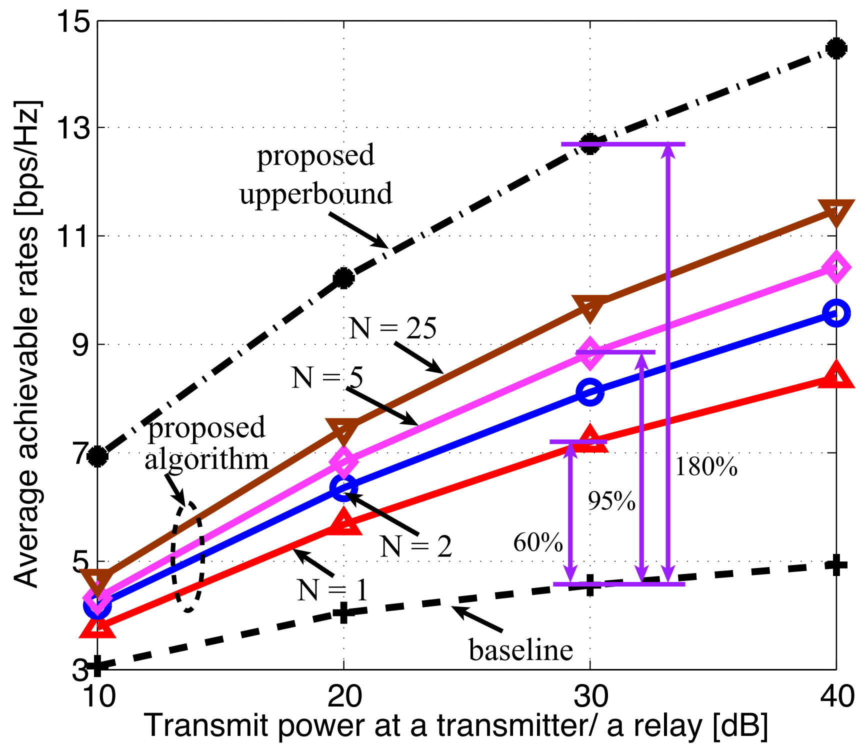

Figure 4.6: Average end-to-end sum-rates of opportunistic solutions of the proposed three-phase algorithm for $N \in \{1, 2, 5, 25\}$ for $(2^3 \times 2^6 \times 2^{12}, 1 \times 1)$.

4.5.3 Opportunistic Solutions

I now investigate the benefits of opportunistic solutions. Fig.4.6 provides the average achievable end-to-end sum-rates of the opportunistic solutions of the proposed three-phase algorithm with the number of initializations $N \in \{1, 2, 5, 25\}$ for the system $(2^3 \times 2^6 \times 2^{12}, 1 \times 1)$. To provide a benchmark, I also show the results for the fixed average normalized second-hop sum-rates, which provides an upper-bound for the solutions. As expected, increasing N improves the end-to-end sum-rates achieved by the proposed algorithm with

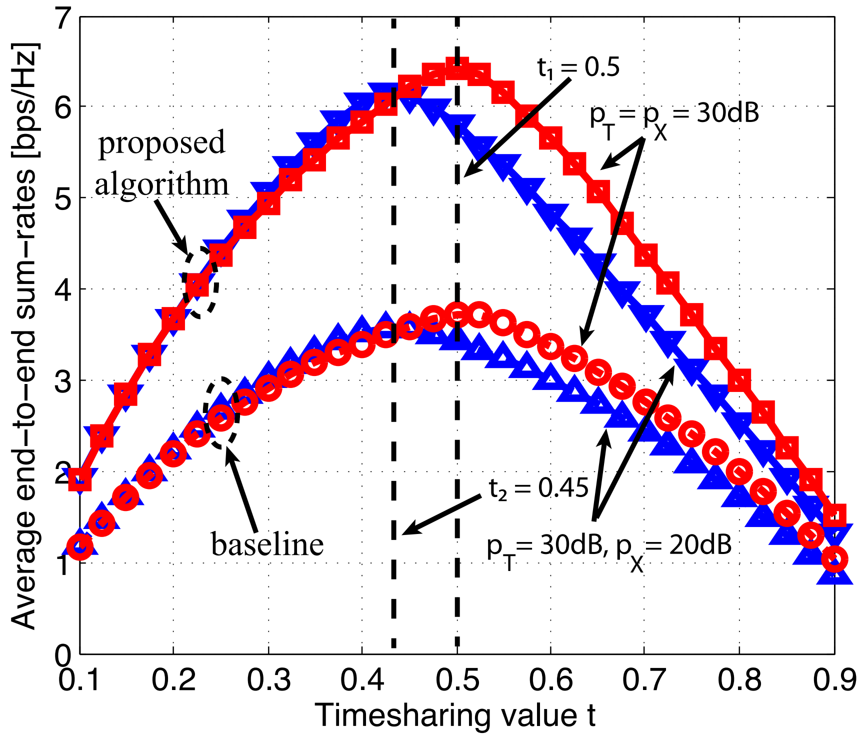

Figure 4.7: Average end-to-end achievable rates as a function of timesharing t for $(2^4 \times 2^8 \times 2^{16}, 1 \times 1)$.

opportunistic implementation. At $p_{\mathrm{T}} = p_{\mathrm{R}} = 30$ dB, the opportunistic solution with $N = 5$ nearly doubles the end-to-end sum-rates when compared to the baseline. Nevertheless, the additional gains obtained by using an additional random initialization in the opportunistic solutions decreases in N. At $p_{\mathrm{T}} = p_{\mathrm{R}} = 30$ dB, the opportunistic solution with $N = 25$ achieves nearly 75% the value of the upper-bound. Nevertheless, note that the opportunistic solutions improve the end-to-end sum-rate performance at the cost of overhead, running time and coordination requirements.

4.5.4 Varying Timesharing Values

Fig. 4.7 presents the average achievable end-to-end sum-rates as functions of the timesharing values for the system $(2^4 \times 2^8 \times 2^{16}, 1 \times 1)$ for the following two cases: i) Case 1 with $p_{\mathrm{T}} = p_{\mathrm{X}} = 30$ dB and ii) Case 2 with $p_{\mathrm{T}} = 30$ dB and $p_{\mathrm{X}} = 20$ dB. I notice that for the proposed algorithm and the baseline in each case has the same optimal timesharing value, $t_1 = 0.5$ for Case 1 and $t_2 = 0.425$ for Case 2. These optimal timesharing values approximately equalize the average normalized sum-rates on two hops. This emphasizes the importance of matching the rates on two hops. In addition, I observe that thanks to the explicit consideration of t for matching the rates on two hops, the proposed algorithm has large gains, between 50% and 70%, over the baseline for $t \in [0.1, 0.9]$.

Chapter 5

Conclusion

I summarize the main results and conclusions of this dissertation in Section 5.1 and discuss possible future research directions in Section 5.2.

5.1 Summary

In this dissertation, I developed transmission strategies for wireless multiple-antenna relay-aided systems. MIMO relay communication is emerging as a viable solution for improving area spectral efficiency in cellular networks. Transmission strategies for jointly configuring the transmitter(s) and relay(s) based on current channel state information are needed to take full advantage of relays to obtain higher data rates and better reliability. Table 5.1 provides a brief comparison of the configuration of multiple-antenna relay systems considered in this dissertation.

In Chapter 2, I proposed dualmode and multimode antenna selection criteria for three-node MIMO AF relay systems with linear ZF receive filters for both the two-hop channel and relay channel to provide reliable transmissions at a guaranteed rate. Although suboptimal, the proposed algorithms could achieve the full diversity gain of the corresponding channel, thus improving

Table 5.1: A summary of the configurations of multiple-antenna relay-assisted systems considered in the previous chapters

Contribution	Relay type	Presence of interference	Multiple users per relay	Design objective
Contribution 1	AF	no	no	VSER minimization
Contribution 2	AF	yes	no	sum-rate maximization
Contribution 3	DF	yes	yes	sum-rate maximization

the diversity performance of plain spatial multiplexing. The dualmode algorithms also provided array gains over single-stream transmission strategies like full selection diversity and limited feedback Grassmannian beamforming. By allowing more options for the number of data streams and stream-to-antenna mappings, the proposed multimode algorithms adapted transmitted signals to current channel conditions better than the other transmission strategies including dualmode algorithms. Thus, the proposed multimode algorithms provided large array gains over the other strategies. In addition, using tools from matrix analysis, I derived closed-form expressions that act like condition numbers of the MIMO AF relay channel. These expressions revealed how the quality of the constituent channels (through their eigenmodes) would affect mode selection. Simulations showed that the eigenmode-based two-hop multimode algorithm works well in an interference-limited multi-cell network and improves the reliability of transmission to cell-edge users.

Chapter 3 focused on interference mitigation strategies for the MIMO AF relay interference channel. I developed three cooperative algorithms for joint configuring the transmitters, relays, and receivers. The first algorithm aimed at minimizing the sum power of interference and enhanced relay noise. Based on a relationship between mutual information and mean square errors, the other two algorithms were able to find stationary points of the end-to-end sum-rate maximization problems with equality (or inequality) power constraints. Simulation results showed that the last two algorithms outperformed the first algorithm at low-to-medium SNR in terms of average end-to-end sum-rates and multiplexing gains. One reason was that the last two algorithms took into account the desired signals as well as noise at the receivers while the first algorithm did not. Nevertheless, they performed worse than the first algorithm at high SNR due to unfairness in achievable rates among users. The multiplexing gains achieved by the proposed algorithms provided lower bounds on the total number of degrees of freedom in MIMO AF relay networks, which remains unknown. Simulations also showed that AF relays enhanced the feasibility of interference alignment at the receivers, leading to higher multiplexing gains than both DF relays and direct transmission.

In Chapter 4, I considered the MIMO DF relay interference broadcast channel. By definition, the end-to-end achievable rate corresponding to a DF relay was equal to the minimum of the achievable (sum) rates on two hops of this relay. Thus, the optimal solutions to the end-to-end sum-rate maximization problem of this system should cause no mismatch between the rates on two hops. The naïve approach of applying existing single-hop interference management strategies separately for two hops of the system caused a two-hop rate mismatch, leading to low end-to-end sum-rates. I developed a three-phase cooperative algorithm for designing the transmit precoders at the transmitters and relays of the MIMO DF relay interference broadcast channel to maximize the achievable end-to-end sum-rates. The timesharing value and second-hop configuration were considered in the design of first-hop transmit precoders. The proposed algorithm could be implemented in a distributed manner and had fast convergence behavior, making it suitable for practical systems. Simulations showed that the proposed algorithm obtained much higher end-to-end sum-rates than the naïve approach.

5.2 Future Work

Link adaptation and interference management for multiple-antenna relay-assisted networks are relatively new areas. In this section, I enumerate several possible directions for future research.

Relay functionality selection The signal processing capabilities at relays play an important role in relay-assisted systems. This dissertation considered only AF and DF relays (see Assumption 2.1, Assumption 3.1, and Assumption 4.1). In the literature, many other signal processing operations at relays have been proposed. For example, the seminal paper [50] proposed a compress-and-forward (CF) relaying technique. Without fully decoding the received signals, CF relays use source coding to compress the signals before retransmitting. This compressed version of the transmitted message provides "side information" for detection at receivers. Another relaying technique, named estimate-and-forward (EF), was introduced in [71]. In principle, EF relays compute and retransmit an estimate of transmitted signals to provide receivers with soft information on the signals. The choice of a specific relaying technique for a given system is somewhat subjective. Such a choice depends on system parameters (e.g., quality of constituent channels, etc.), system design objectives (e.g., sum-rate maximization, error-rate minimization, etc.) and system design constraints (e.g., computational complexity). Different relay techniques may be optimal in different scenarios. For example, in a three-node relay channel, if the SNR of the relay-to-receiver channel is high enough so that the receiver can perfectly obtain the compressed signals from the relay, then CF can be capacity-optimal [112]. As another example, EF is optimal in terms of minimizing bit error rate or maximizing SNR at the receiver in the case of discrete input (such as binary phase-shift keying). As discussed in Section 1.2, the choice between only AF and DF relays is complicated. The presence of interference makes such choices more challenging. Thus, future work should investigate various relaying techniques in different scenarios to obtain insights into relay functionality selection, especially in the presence of interference.

Link adaptation Chapter 2 presented adaptive antenna selection algorithms for improving the reliability of fixed-rate transmission (see Assumption 2.9 and Assumption 2.5). Future work can develop adaptive algorithms for maximizing end-to-end sum-rate, which is arguably a more important design objective in cellular systems, or other objectives. The impact of out-of-cell interference should also be accounted for in the development of link adaptation strategies for multiple-antenna relay-assisted cellular systems. Future work should extend the results to develop adaptive precoding algorithms for the MIMO relay channel. In linear precoded spatial multiplexing systems, the transmitted data vector is premultiplied by a precoding matrix that is designed based on some form of channel information. Antenna selection is a special type of linear precoding where the precoding matrices are constrained to columns of the relevant identity matrices. The use of more general linear precoding matrices by relaxing the constraints improves reliability and data rates at the cost of higher complexity.

Another possible research direction is to develop online supervised learning link adaption algorithms for relay networks. Applying tools from statistical machine learning, prior work in [51, 52] develops an online supervised learning framework for link adaptation in the wireless point-to-point channel. The key is that by defining appropriate feature space, any changes in the wireless system operation and wireless propagation medium can be captured in the related data observations. Therefore, the proposed algorithm allows these wireless systems to select adaptively a plurality of parameters at different layers to optimize network throughput while satisfying certain reliability constraints. With the framework, what remains is to find appropriate feature space for extracting information on wireless system parameters. Future work can extend the prior algorithms to those for the MIMO relay channel. Note that the concepts of condition numbers proposed in Chapter 2 can be used as feature space for extracting information on the spatial characteristics of the corresponding MIMO relay channel. Nevertheless, future work could find better feature space.

Robust algorithms in the presence of CSI uncertainty Previous chapters of this dissertation assumed perfect CSI, i.e., all the coefficients of constituent point-to-point channels, both desired and interfering, were known instantaneously and perfectly where needed (see Assumption 2.4, Assumption 3.8, and Assumption 4.7). Nevertheless, obtaining this perfect CSI condition may be difficult due to the cost of channel coefficient estimation (e.g., when network size is large) and the impairments of feedback channels used to convey CSI. Recall that simulations in Chapter 2 showed that the VSER performance of the proposed multimode algorithms for the MIMO AF relay channel degrades with the increase of feedback delay. Thus, future work should investigate and quantify the impact of various levels of CSI uncertainly on the proposed algorithms. Future work can also focus on developing algorithms that provide robustness against CSI uncertainty. Such algorithms can be evaluated in relative comparison with the benchmarks provided by the counterpart algorithms proposed in this dissertation. Due to its importance in practical scenarios, robust signal processing for wireless communications has been an important research area, e.g., see [53] and references therein. Many robust algorithms have been proposed for designing transceivers in the three-node AF relay systems [37, 219], in the MIMO interference channel [44, 182], and the MIMO AF relay interference channel [38]. Note that prior work in [38] aims at either minimizing sum power while fulfilling the SINR requirements or maximizing the minimum of the SINRs subject to transmit power constraints, but not at maximizing end-to-end sum-rates.

Theoretical investigation of multiple-antenna relay-aided interference alignment Interference alignment is not always feasible. The prior work in [166, 225] considers the symmetric MIMO single-hopinterference channel with constant channel coefficients. The maximum total number of independent data streams that can be sent over this MIMO interference channel using linear transceivers is upper-bounded by the total number of antennas at each transmitter-receiver pair. Such theoretical results provide important insights into the feasibility of interference alignment in the MIMO single-hop interference channel. Thus, similar theoretical results for the multiple-antenna relay interference channel are desired. Nevertheless, there has been little theoretical investigation of general MIMO relay-aided interference alignment. Many of the existing results are only for the extreme (either very small or very large) systems [74, 140] or simplified systems [145]. In Chapter 3, I assume that the considered configurations make the end-to-end interference alignment feasible (see Assumption 3.14). Future work could investigate and derive the feasibility conditions and achievable total degrees of freedom for multiple-antenna relay-assisted interference alignment for both AF and DF relay networks. A possible approach is to extend the results for the MIMO single-hop interference channel in [166, 225] to account for special features of relay communication such as multi-hop transmission and relay signal processing operation. Such extensions are not straightforward.

Mixed heterogeneous networks: Note that this dissertation focused only on one type of low-power node in heterogeneous networks, i.e., relays. Much prior work also considers only one type of low-power nodes, e.g., [85, 87, 137] for distributed antenna systems and [39, 104, 151] for femtocell networks. Also, I assume that there is no uncoordinated interference (see Assumption 2.12, Assumption 3.10, and Assumption 4.11). As discussed in Section 1.4, however, future heterogeneous networks will have a multi-tier architecture with a mixture of relays, femtocells, pico-cells and distributed antenna systems. All these low-power nodes may cover the same geographical area using the same radio resources as macro-cells. Their transmissions are likely to interfere with each other, leading to multi-tier interference. This motivates the need for understanding the interactions among various wireless access technologies in future heterogeneous networks. This is challenging because each technology operates under different deployment assumptions and creates a different interference environment. Future work can analyze the performance of such mixed heterogeneous networks. A recent result [82] provides a simplified interference model for such mixed heterogenous networks. The key idea is to analyze the performance of a typical cell assuming the interferers outside the cell of interest are distributed according to a superposition of marked Poisson point processes. Each marked Poisson point process represents a set of low-power nodes of the same type. This framework may allow for analyzing and evaluating various complex multi-tier interference scenarios. Future work can also develop interference management strategies for such mixed heterogeneous networks. For example, the interference management strategies for relay-only networks proposed in Chapter 3 and Chapter 4 can be extended to those for mixed networks that take into account the cross-tier interference due to femtocells into relay design.

Appendices

Appendix A

Rank-Reduction Procedure

For completeness, this appendix summarizes the rank-reduction procedure to construct a rank-one global optimum of a separable semidefinite programming (SDP) from its general-rank global optimum. The formulation of the separable SDP is given by

$$\begin{aligned}(P0):\ \min_{\mathbf{X}} \quad & \mathrm{tr}(\mathbf{A}\mathbf{X}) \\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{B}_m\mathbf{X}) = b_m, m = 1, \cdots, M \\ & \mathbf{X} \succeq \mathbf{0},\end{aligned} \tag{A.1}$$

where $M \leq 3$. Since $(P0)$ is convex, its general-rank global optimum can be found efficiently and to any arbitrary accuracy by using available software packages for convex optimization like CVX [77]. Let $\mathbf{X}_0$ denote the resulting general-rank global optimum of $(P0)$. A rank-one global optimum of $(P0)$ can be constructed from $\mathbf{X}_0$ via the following rank-reduction procedure, which is a version of Algorithm 1 in [95] with my notation. The rank-one procedure is as follows

- Initialization $\mathbf{Y} = \mathbf{X}_0$
- Evaluate $r = \mathrm{rank}(\mathbf{Y})$

- ***While*** $r > \sqrt{M}$
  - Decompose $\mathbf{Y} = \mathbf{V}\mathbf{V}^*$
  - Find a nonzero solution $\mathbf{Z}$ of the system of linear equations

$$\operatorname{tr}(\mathbf{V}^*\mathbf{B}_m\mathbf{V}\mathbf{Z}) = 0, m = 1, \cdots, M, \tag{A.2}$$

    where $\mathbf{Z} \in \mathbb{C}^{r\times r}$ is a Hermitian matrix
  - Evaluate the eigenvalues $\delta_1, \cdots, \delta_r$ of $\mathbf{Z}$
  - Determine k_0 such that

$$\delta_{k_0} = \max\{|\delta_k| : \ 1 \leq k \leq r\} \tag{A.3}$$

  - Compute $\mathbf{Y} = \mathbf{V}(\mathbf{I}_r - (1/\delta_{k_0})\mathbf{Z})\mathbf{V}^*$
  - Evaluate $r = \operatorname{rank}(\mathbf{Y})$
- ***End while***

The output of the rank-reduction procedure is $\mathbf{Y}$, which is a global optimum of $(P0)$. The rank of $\mathbf{Y}$ satisfies the following condition $\operatorname{rank}(\mathbf{Y}) \leq \sqrt{M} < 2$ since $M \leq 3$. Thus, $\mathbf{Y}$ is a rank-one global optimum of $(P0)$.

Bibliography

[1] The mobile internet report. Technical report, Morgan Stanley, December 2009, [Online] Available: http://www.morganstanley.com/institutional/techresearch/pdfs/mobile_internet_report.pdf.

[2] ICT facts and figures 2011. Technical report, International Telecommunication Union, October 2011, [Online] Available: http://www.itu.int/ITU-D/ict/facts/2011.

[3] Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016. Technical report, Cisco, February 2012, [Online] Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11520862.pdf.

[4] 3GPP TR 36.806. Evolved universal terrestrial radio access (E-UTRA); relay architectures for E-UTRA (LTE-Advanced). [Online] Available: http://www.3gpp.org/ftp/Specs/html-info/36806.htm.

[5] 3GPP TR 36.814. Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspect. [Online] Available: http://www.3gpp.org/ftp/Specs/html-info/36814.htm.

[6] 3GPP TR 36.826. Evolved universal terrestrial radio access (E-UTRA); relay radio transmission and reception. [Online] Available:

http://www.3gpp.org/ftp/Specs/html-info/36826.htm.

[7] T. Abe, H. Shi, T. Asai, and H. Yoshino. Relay techniques for MIMO wireless networks with multiple source and destination pairs. *EURASIP Journal on Wireless Communications and Networking*, 2006, 2006.

[8] S. Akoum and R. W. Heath, Jr. Limited feedback for temporally correlated MIMO channels with other cell interference. *IEEE Transactions on Signal Processing*, 58(10):5219–5232, October 2010.

[9] S. H. Ali and V. C. M. Leung. Dynamic frequency allocation in fractional frequency reused OFDMA networks. *IEEE Transactions on Wireless Communications*, 8(8):4286–4295, August 2009.

[10] W. Ali, Y. Huang, and S. Zhang. New results on Hermitian matrix rank-one decomposition. *Mathematical Programming, Series A*, 120(2), 2009.

[11] M.-S. Alouini and A. J. Goldsmith. Area spectral efficiency of cellular mobile radio systems. *IEEE Transactions on Vehicular Technology*, 48(4):1047–1066, July 1999.

[12] M.-S. Alouini and M. K. Simon. Performance analysis of coherent equal gain combining over Nakagami-m fading channels. *IEEE Transactions on Vehicular Technology*, 50(6):1449–1463, November 2001.

[13] J. G. Andrews, F. Baccelli, and R. K. Ganti. A tractable approach to coverage and rate in cellular networks. *IEEE Transactions on Communications*, 59(11):3122–3134, November 2011.

[14] J. G. Andrews, W. Choi, and R. W. Heath, Jr. Overcoming interference in spatial multiplexing MIMO cellular networks. *IEEE Wireless Communications*, 14(6):95–104, December 2007.

[15] K. E. Baddour and N. C. Beaulieu. Autoregressive modeling for fading channel simulation. *IEEE Transactions on Wireless Communications*, 4(4):1650–1662, July 2005.

[16] Y. Bae, S. Jung, and J. Lee. Capacity comparison of orthogonal and non-orthogonal cooperative relay systems. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, April 2009.

[17] A. Barbieri, P. Gaal, S. Geirhofer, T. Ji, D. Malladi, Y. Wei, and F. Xue. Coordinate downlink multi-point communications in heterogeneous 4G cellular networks. In *Proc. of IEEE Workshop on Information Theory Applications*, February 2012.

[18] T. Beniero, S. Redana, J. Hamalainen, and B. Raaf. Effect of relaying on coverage in 3GPP LTE-Advanced. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, Barcelona, Spain, April 2009.

[19] S. Berger, M. Kuhn, A. Wittneben, T. Unger, and A. Klein. Recent advances in amplify-and-forward two-hop relaying. *IEEE Communication*

Magazine, 47(7):50–56, July 2009.

[20] R. Bhagavatula. *Limited Feedback for Multicell Cooperative Systems.* PhD thesis, The University of Texas at Austin, December 2010.

[21] R. Bhagavatula and R. W. Heath, Jr. Adaptive bit partitioning for multicell intercell interference cancellation with delayed limited feedback. *IEEE Transactions on Signal Processing*, 59(8):3824–3836, April 2010.

[22] R. Bhagavatula and R. W. Heath, Jr. Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems. *IEEE Transactions on Signal Processing*, 59(2):800–811, November 2010.

[23] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman. A simple cooperative diversity method based on network path selection. *IEEE Journal on Selected Areas in Communications*, 24(3):659–672, March 2006.

[24] R. S. Blum. MIMO capacity with interference. *IEEE Journal on Selected Areas in Communications*, 21(5), June 2003.

[25] F. Boccardi and H. Huang. Limited downlink network coordination in cellular networks. In *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communcations*, pages 1–5, September 2007.

[26] F. Boccardi, H. Huang, and A. Alexiou. Network MIMO with reduced backhaul requirements by MAC coordination. In *Proc. of IEEE Asilo-*

mar Conference on Signals, Systems and Computers, pages 1125–1129, October 2008.

[27] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom of the K-user interference channel. *IEEE Transactions on Information Theory*, 54(8):3425–3441, August 2008.

[28] V. R. Cadambe and S. A. Jafar. Degrees of freedom of wireless networks with relays, feedback, cooperation, and full duplex operation. *IEEE Transactions on Information Theory*, 55(5):2334–2344, May 2009.

[29] B. Can, H. Yomo, and E. De Varvalho. Link adaptation and selection method for OFDM based wireless relay networks. *Journal on Comunications and Networking*, 9(2):118–127, June 2007.

[30] Y. Cao and B. Chen. Capacity bounds for two-hop interference networks. In *Proc. of Allerton Conference on Communication, Control, and Computing*, pages 272–279, Montecillo, IL, September-October 2009.

[31] S. Catreux, P. F. Driessen, and L. J. Greenstein. Simulation results for an interference-limited multiple-input multiple-output cellular system. *IEEE Communications Letters*, 4(11):334–336, November 2000.

[32] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, Jr. Adaptive modulation and MIMO coding for broadband wireless data networks. *IEEE Communications Magazine*, 40(6):108–115, June 2002.

[33] A. Chaaban and A. Sezgin. From non-causality to causality: degrees of freedom of the interference relay channel. In *Proc. of Conference on Information Sciences Systems*, pages 1–6, Princeton, NJ, March 2010.

[34] C. B. Chae, S. Kim, and R. W. Heath, Jr. Network coordinated beamforming for cell-boundary users: Linear and non-linear approaches. *IEEE Journal on Selected Topics in Signal Processing*, 3(6):1094–1105, December 2009.

[35] C.-B. Chae, T. Tang, R. W. Heath, Jr., and S. Cho. MIMO relaying with linear processing for multiuser transmission in fixed relay networks. *IEEE Transactions on Signal Processing*, 56(2):727–738, February 2008.

[36] A. Chakrabarti, A. Sabharwal, and B. Aazhang. Sensitivity of achievable rates for half-duplex relay channel. In *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications*, pages 970–974, New York City, NY, June 2005.

[37] B. K. Chalise and L. Vandendorpe. Joint linear processing for an amplify-and-forward MIMO relay channel with imperfect channel state information. *EURASIP Journal on Advances in Signal Processing*, 2010, 2010.

[38] B. K. Chalise and L. Vandendorpe. Optimization of MIMO relays for multipoint-to-multipoint communications: Nonrobust and robust designs. *IEEE Transactions on Signal Processing*, 58(12):6355–6368, December 2010.

[39] V. Chandrasekhar, J. G. Andrews, and A. Gatherer. Femtocell networks: a survey. *IEEE Communications Magazine*, 46(9):59–67, September 2008.

[40] H. Chen, A. B. Gershman, and S. Shahbazpanahi. Distributed peer-to-peer beamforming for multiuser relay networks. In *Proc. of IEEE International Conference on Acoustics Speech Signal Processing*, pages 2265–2268, Taipei, Taiwan, April 2009.

[41] S. Chen and R. S. Cheng. Achieve the degrees of freedom of K-user MIMO interference channel with a MIMO relay. In *Proc. of IEEE Global Communications Conference*, pages 1–5, Miami, FL, December 2010.

[42] Z. Chen, H. Liu, and W. Wang. Optimal transmit strategy of a two-hop decode-and-forward MIMO relay system with mean and covariance feedback. *IEEE Communications Magazine*, 14(6):530–532, june 2010.

[43] M. Chiani, M. Z. Win, and H. Shin. MIMO networks: The effects of interference. *IEEE Transactions on Information Theory*, 56(1):336–349, January 2010.

[44] E. Chiu, V. K. N. Lau, H. Huang, T. Wu, and S. Liu. Robust transceiver design for K-pairs quasi-static MIMO interference channels via semi-definite relaxation. *IEEE Transactions on Wireless Communications*, 9(12):3762–3769, December 2010.

[45] W. Choi and J. G. Andrews. The capacity gain from intercell scheduling in multi-antenna systems. *IEEE Transactions on Wireless Communications*, 7(2):714–725, February 2008.

[46] A. Chopra and B. L. Evans. Joint statistics of radio frequency interference in multi-antenna receivers. *to appear IEEE Transactions on Signal Processing*, 2012.

[47] Aditya Chopra. *Modeling and Mitigation of Interference in Wireless Receivers with Multiple Antennae*. PhD thesis, The University of Texas at Austin, Austin, TX, USA, December 2011.

[48] S. S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. *IEEE Transactions on Wireless Communications*, 7(12):4792–4799, December 2008.

[49] H. Claussen and L. G. Samuel. The future of small cell networks. *IEEE COMSOC MMTC E-letter*, 5(5), September 2010.

[50] T. Cover and A. El Gamal. Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25(5):572–584, September 1979.

[51] R. C. Daniels, C. Caramanis, and R. W. Heath, Jr. A supervised learning approach to adaptation in practical MIMO-OFDM wireless sys-

tems. In *Proc. of IEEE Global Communications Conference*, pages 1 –5, November-December 2008.

[52] R. C. Daniels and R. W. Heath, Jr. An online learning framework for link adaptation in wireless networks. In *Proc. of IEEE Workshop on Information Theory and Applications*, pages 138–140, February 2009.

[53] F. A. Dietrich. *Robust Signal Processing for Wireless Communications.* Springer, 2008.

[54] K. Doppler, C. Wijting, and K. Valkealahti. On the benefits of relays in a metropolitan area network. In *Proc. of IEEE Vehicular Technology Conference*, pages 2301–2305, May 2008.

[55] E. H. Drucker. Development and application of a cellular repeater. In *Proc. of IEEE Vehicular Technology Conference*, pages 321–325, January 1988.

[56] A. Ekbal and J. M. Cioffi. Distributed transmit beamforming in cellular networks - a convex optimization perspective. In *Proc. of IEEE International Conference on Communications*, pages 2690–2694, May 2005.

[57] C. Esli and A. Wittneben. A hierarchical AF protocol for distributed orthogonalization in multiuser relay networks. *IEEE Transactions on Vehicular Technology*, 59(8):3902–3916, October 2010.

[58] Y. Fan and J. Thompson. MIMO configurations for relay channels: Theory and practice. *IEEE Transactions on Wireless Communications*, 6(5):1774 –1786, May 2007.

[59] Z. Fang, Y. Hua, and J. C. Koshy. Joint source and relay optimization for a non-regenerative MIMO relay. In *Proc. of IEEE Workshop on Sensor Array Multichannel Processing*, pages 239–243, Waltham, MA, July 2006.

[60] S. Fazeli-Dehkordy, S. Shahbazpanahi, and S. Gazor. Multiple peer-to-peer communications using a network of relays. *IEEE Transactions on Signal Processing*, 57(8):3053–3062, August 2009.

[61] G. J. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multiple antennas. *Bell Labs Technical Journal*, 1(2):41–59, 1996.

[62] G. J. Foschini and M. J. Gans. On the limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, (6):311–335, 1998.

[63] G. J. Foschini and Z. Miljanic. A simple distributed autonomous power control algorithm and its convergence. *IEEE Transactions on Vehicular Technology*, 42(4):641–646, November 1993.

[64] A. Ghasemi, A. S. Motahari, and A. K. Khandani. Interference alignment for the K user MIMO interference channel. In *Proc. of IEEE In-*

ternational Symposium on Information Theory, pages 360–364, Austin, TX, June 2010.

[65] A. Ghosh, J. G. Andrews, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, P. Xia, H. S. Jo, H. S. Dhillo, and T. D. Novlan. Heterogeneous cellular networks: From theory to practice. *IEEE Communications Magazine*, June 2012.

[66] H. Ghozlan, Y. Mohasseb, G. Kramer, and H. El Gamal. The MIMO wireless switch: Relaying can increase the multiplexing gain. In *Proc. of IEEE International Symposium on Information Theory*, pages 1448–1452, July 2009.

[67] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[68] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[69] K. Gomadam, V. R. Cadambe, and S. A. Jafar. Approaching the capacity of wireless networks through distributed interference alignment. In *Proc. of IEEE Global Communications Conference*, pages 1–6, New Orleans, LA, November-December 2008.

[70] K. Gomadam, V. R. Cadambe, and S. A. Jafar. A distributed numerical approach to interference alignment and applications to wireless interfer-

ence networks. *IEEE Transactions on Information Theory*, 57(6):3309–3322, June 2011.

[71] K. S. Gomadam and S. A. Jafar. Optimal relay functionality for SNR maximization in memoryless relay networks. *IEEE Journal on Selected Areas in Communications*, 25(2):390–401, February 2007.

[72] K. S. Gomadam and S. A. Jafar. Duality of MIMO multiple access channel and broadcast channel with amplify-and-forward relays. *IEEE Transactions on Communications*, 58(1):211–217, January 2010.

[73] T. Gou and S. A. Jafar. Degrees of freedom of the K user M x N MIMO interference channel. *IEEE Transactions on Information Theory*, 56(12):6040–6057, December 2010.

[74] T. Gou, S. A. Jafar, S.-W. Jeon, and S.-Y. Chung. Aligned interference neutralization and degrees of freedom of the 2x2x2 interference channel. *submitted to IEEE Transactions on Information Theory*, December 2010.

[75] S. Govindasamy, D. W. Bliss, and D. H. Staelin. Spectral-efficiency of multi-antenna links in ad-hoc wireless networks with limited Tx CSI. In *Proc. of IEEE Asilomar Conference on Signals, Systems and Computers*, pages 1714–1718, November 2009.

[76] K. Gracie and M.-H. Hamon. Turbo and turbo-like codes: Principles and applications in telecommunications. *Proceedings of the IEEE*, 95(6):1228–1254, June 2007.

[77] M. Grant and S. Boyd. CVX: MATLAB software for disciplined convex programming. `http://cvxr.com/cvx`, April 2011.

[78] W. Guan and H. Luo. Joint MMSE transceiver design in non-regenerative MIMO relay systems. *IEEE Communications Letters*, 12(7):517–519, July 2008.

[79] W. Guan, H. Luo, and W. Chen. Linear relaying scheme for MIMO relay system with QoS requirements. *IEEE Signal Processing Letters*, 15:697–700, 2008.

[80] M. Hajiaghayi, M. Dong, and B. Liang. Using limited feedback in power allocation design for a two-hop relay OFDM system. In *Proc. of IEEE International Conference on Communications*, pages 1–6, June 2009.

[81] J. F. Hart. *Computer Approximations*. Krieger Pub. Co., 1978.

[82] R. W. Heath, Jr. and M. Kountouris. Modeling heterogeneous network interference. In *Prof. of IEEE Workshop on Information Theory and Applications*, San Diego, CA, Feburary 2012.

[83] R. W. Heath, Jr. and D. J. Love. Multimode antenna selection for spatial multiplexing systems with linear receivers. *IEEE Transactions on Signal Processing*, 53(8):3042–3056, August 2005.

[84] R. W. Heath, Jr. and A. J. Paulraj. Switching between diversity and multiplexing in MIMO systems. *IEEE Transactions on Communications*, 53(6):962–968, June 2005.

[85] R. W. Heath, Jr., S. W. Peters, Y. Wang, and J. Zhang. A current perspective on distributed antenna systems. *submitted to IEEE Communications Magazine*, December 2011.

[86] R. W. Heath, Jr., S. Sandhu, and A. J. Paulraj. Antenna selection for spatial multiplexing systems with linear receivers. *IEEE Communications Letters*, 5(4):142–144, April 2001.

[87] R. W. Heath, Jr., T. Wu, Y.-H. Kwon, and A. C. K. Soong. Multiuser MIMO in distributed antenna systems with out-of-cell interference. *IEEE Transactions on Signal Processing*, 59(10):4885–4899, October 2011.

[88] Z. K. M. Ho and D. Gesbert. Balancing egoism and altruism on interference channel: The MIMO case. In *Proc. of IEEE International Conference on Communications*, pages 1–5, Cape Town, South Africa, May 2010.

[89] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[90] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Press Syndicate of the University of Cambridge, 1991.

[91] A. Host-Madsen and A. Nosratinia. The multiplexing gain of wireless networks. In *Proc. of IEEE International Symposium on Information Theory*, pages 2065–2069, September 2005.

[92] A. Host-Madsen and J. Zhang. Capacity bounds and power allocation for wireless relay channels. *IEEE Transactions on Information Theory*, 51(6):2020–2040, Jun. 2005.

[93] H. Hu, H. Yanikomeroglu, D. D. Falconer, and S. Periyalwar. Range extension without capacity penalty in cellular networks with digital fixed relays. In *Proc. of IEEE Global Communications Conference*, pages 3053–3057, Dallas, TX, November-December 2004.

[94] J. Huang, R. A. Berry, and M. L. Honig. Distributed interference compensation for wireless networks. *IEEE Journal on Selected Areas in Communications*, 24(5):1074–1084, May 2006.

[95] Y. Huang and D. P. Palomar. Rank-constrained separable semidefinite programming with applications to optimal beamforming. *IEEE Transactions on Signal Processing*, 58(2):664–678, February 2010.

[96] Y. Huang, Y. Yang, M. Bengtsson, and B. Ottersten. A limited feedback joint precoding for amplify-and-forward relaying. *IEEE Transactions on Signal Processing*, 58(3):1347–1357, March 2010.

[97] IEEE 802.16j. Standard for local and metropolitan area networks - Part 16: Air interface for broadband wireless access systems - Amendment 1: Multiple relay spefication. [Online] Available: http://standards.ieee.org/getieee802/download/802.16j-2009.pdf.

[98] IEEE 802.16m Broadband Wireless Access Working Group. IEEE 802.16m system requirement. [Online] Available: http://standards.ieee.org/getieee802/download/802.16m-2011.pdf.

[99] R. Irmer and F. Diehm. On coverage and capacity of relaying in LTE-Advanced in example deployments. In *Proc. of IEEE Conference on Personal, Indoor, and Mobile Radio Communications*, pages 1–5, September 2008.

[100] S. A. Jafar. *Interference Alignment - A new Look at Signal Dimensions in a Communication Network*, volume 7, pages 1–134. Now Publishers Inc., 2011.

[101] S. A. Jafar, G. J. Foschini, and A. J. Goldsmith. PhantomNet: exploring optimal multicellular multiple antenna systems. In *Proc. of IEEE Vehicular Technology Conference*, volume 1, pages 261–265, June 2002.

[102] S. A. Jafar, K. S. Gomadam, and C. Huang. Duality and rate optimization for multiple access and broadcast channels with amplify-and-forward relays. *IEEE Transactions on Information Theory*, 53(10):3350–3370, October 2007.

[103] S.-W. Jeon, S.-Y. Chung, and S. A. Jafar. Degrees of freedom region of a class of multisource Gaussian relay networks. *IEEE Transactions on Information Theory*, 57(5):3032–3044, May 2011.

[104] H.-C. Jo, C. Mun, J. Moon, and J.-G. Yook. Interference mitigation using uplink power control for two-tier femtocell networks. *IEEE Transactions on Wireless Communications*, 8(10):4906–4910, October 2009.

[105] E. A. Jorswieck, E. G. Larsson, and D. Danev. Complete characterization of the Pareto boundary for the MISO interference channel. *IEEE Transactions on Signal Processing*, 56(10):5292–5296, October 2008.

[106] J. Joung and A. H. Sayed. Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems. *IEEE Transactions on Signal Processing*, 58(3):1833–1846, March 2010.

[107] N. Khaled, S. Thoen, and L. Deneire. Optimizing the joint transmit and receive MMSE design using mode selection. *IEEE Transactions on Communications*, 53(4):730–737, April 2005.

[108] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi. LTE-Advanced: Heterogeneous networks. In *European Wireless Conference*, pages 978–982, Lucca, Italy, April 2010.

[109] B. Khoshnevis, W. Yu, and R. Adve. Grassmannian beamforming for MIMO amplify-and-forward relaying. *IEEE Journal on Selected Areas in Communications*, 26(8):1397–1407, October 2008.

[110] K. Ko and C.-L. Lin. *Minimax and Applications*, chapter On the complexity of min-max optimization problems and their approximation, pages 219–239. Kluwer, 1995.

[111] K. Kobayashi, T. Ohtsuki, and T. Kaneko. MMSE precoder with mode selection for MIMO systems. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, September 2006.

[112] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51(9):3037–3063, September 2005.

[113] J. N. Laneman. *Cooperative diversity in wireless networks: algorithms and architectures*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, August 2002.

[114] J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50(12):3062–3080, December 2004.

[115] E. Larsson and E. Jorswieck. Competition versus cooperation on the MISO interference channel. *IEEE Journal on Selected Areas in Communications*, 26(7):1059–1069, September 2008.

[116] B. Lee, H. Je, O.-S. Shin, and K. Lee. A novel uplink MIMO transmission scheme in a multicell environment. *IEEE Transactions on Wireless Communications*, 8(10):4981–4987, October 2009.

[117] I.-H. Lee and D. Kim. Coverage extension and power allocation in dual-hop space-time transmission with multiple antennas in each node.

IEEE Transactions on Vehicular Technology, 56(6):3524 –3532, November 2007.

[118] J. Lee, J.-K. Han, and J. Zhang. MIMO technologies in 3GPP LTE and LTE-Advanced. *EURASIP Journal on Wireless Communications and Networking*, 2009, 2009.

[119] K.-J. Lee, H. Sung, E. Park, and I. Lee. Joint optimization for one and two-way MIMO AF multiple-relay systems. *IEEE Transactions on Wireless Communications*, 9(12):3671–3681, December 2010.

[120] N. Lee and S. A. Jafar. Aligned interference neutralization and the degrees of freedom of the 2 user interference channel with instantaneous relay. *submitted to IEEE Transactions on Information Theory*, 2011.

[121] Y.-W. Liang and R. Schober. Amplify-and-forward multi-antenna beamforming with joint source-relay power constraint. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, Taipei, Taiwan, September 2010.

[122] Y. Liu and A. P. Petropulu. Cooperative beamforming in multi-source multi-destination relay systems with SINR constraints. In *Proc. of IEEE International Conference on Acoustics Speech Signal Processing*, pages 2870–2873, Dallas, TX, March 2010.

[123] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo. On the complexity of optimal coordinated downlink beamforming. In *Proc. of IEEE International*

Conference on Acoustics Speech Signal Processing, Dallas, TX, March 2010.

[124] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo. Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms. *IEEE Transactions on Signal Processing*, 59(3):1142–1157, Mar. 2011.

[125] Z. Liu, X. Wang, H. Zhang, and Z. Song. Outage analysis of multi-antenna DF relay systems with finite feedbacks over Nakagami-m fading channels. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, May 2010.

[126] C. K. Lo, S. Vishwanath, and R. W. Heath, Jr. Rate bounds for MIMO relay channels. *Journal of Communications and Networks*, 10(2):194–203, June 2008.

[127] K. Loa, C.-C. Wu, S.-T. Sheu, Y. Yuan, M. Chion, D. Huo, and L. Xu. IMT-Advanced relay standards. *IEEE Communications Magazine*, 48(8):40–48, August 2010.

[128] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Q. S. Quek, and J. Zhang. Enhanced intercell interference coordination challenges in heterogeneous networks. *IEEE Wireless Communications*, 18(3):22–30, June 2011.

[129] D. J. Love. Grassmannian subspace packing. [Online] Available: http://cobweb.ecn.purdue.edu/˜djlove/grass.html.

[130] D. J. Love and R. W. Heath, Jr. Multimode precoding for MIMO wireless systems. *IEEE Transactions on Signal Processing*, 53(10):3674–3687, October 2005.

[131] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine*, 27(3):20–34, May 2010.

[132] H. Lutkepohl. *Handbook of Matrices*. John Wiley and Sons, Ltd., 1996.

[133] S. Ma, Y. Yang, and H. Sharif. Distributed MIMO technologies in cooperative wireless networks. *IEEE Communications Magazine*, 49(5):78–82, May 2011.

[134] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji. Cell association and interference coordination in heterogeneous LTE-A cellular networks. *IEEE Journal on Selected Areas in Communications*, 28(9):1479–1489, December 2010.

[135] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani. Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis. *IEEE Transactions on Information Theory*, 54(8):3457–3470, August 2008.

[136] L. Mailaender. Adapting the number of transmitting antennas in mobile networked MIMO (MNM). In *Proc. of IEEE Military Communications Conference*, volume 4, pages 2019–2023, October 2005.

[137] K. Maruta, T. Maruyama, A. Ohta, and M. Nakatsugawa. Inter-cluster interference canceller for multiuser MIMO distributed antenna systems. In *Proc. of IEEE Personal Indoor Mobile Radio Commun.*, pages 3079–3083, September 2009.

[138] J. Melzer and J. Slevinsky. Securing WLAN offload of cellular networks using subscriber residential access gateways. In *IEEE International Conference on Consumer Electronics*, pages 905–906, January 2011.

[139] A. F. Molisch and M. Z. Win. MIMO systems with antenna selection. *IEEE Microwave Magazine*, 5(1):46–56, March 2004.

[140] V. I. Morgenshtern and H. Bolcskei. Crystallization in large wireless networks. *IEEE Transactions on Information Theory*, 53(10):3319–3349, October 2007.

[141] O. Munoz-Medina, J. Vidal, and A. Agustin. Linear transceiver design in nonregenerative relays with channel state information. *IEEE Transactions on Signal Processing*, 55(6):2593–2604, June 2007.

[142] R. U. Nabar, H. Bolcskei, and F. W. Kneubuhler. Fading relay channels: performance limits and space-time signal design. *IEEE Journal on Selected Areas in Communications*, 22(6):1099–1109, August 2004.

[143] T. C.-Y. Ng and W. Yu. Joint optimization of relay strategies and resource allocations in cooperative cellular networks. *IEEE Journal on Selected Areas in Communications*, 25(2):328–339, Feb. 2007.

[144] D. H. N. Nguyen, H. H. Nguyen, and T. T. Pham. Distributed beamforming in multiuser multi-relay networks with guaranteed QoS. In *Proc. of IEEE Global Communications Conference*, pages 1–6, December 2009.

[145] H. Ning, C. Ling, and K. K. Leung. Relay-aided interference alignment: Feasibility conditions and algorithm. In *Proc. of IEEE International Symposium on Information Theory*, pages 390–394, Austin, TX, June 2010.

[146] B. Nourani, S. A. Motahari, and A. K. Khandani. Relay-aided interference alignment for the quasi-static X channel. In *Proc. of IEEE International Symposium on Information Theory*, pages 1764–1768, June 2009.

[147] B. Nourani, S. A. Motahari, and A. K. Khandani. Relay-aided interference alignment for the quasi-static interference channel. In *Proc. of IEEE International Symposium on Information Theory*, pages 405–409, Austin, TX, June 2010.

[148] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Communications Magazine*, 42(9):80–89, September 2004.

[149] A. Y. Panah. *Relay-Aided Communications with Partial Channel State*

Information. PhD thesis, The University of Texas at Austin, Austin, TX, August 2011.

[150] A. Y. Panah, K. T. Truong, S. W. Peters, and R. W. Heath, Jr. Interference management schemes for the shared relay concept. *EURASIP Journal on Advances in Signal Processing*, 2011, Article ID 269817, 14 pages, 2011.

[151] F. Pantisano, M. Bennis, W. Saad, and M. Debbah. Cooperative interference alignment in femtocell networks. In *Proc. of IEEE Global Communications Conference*, pages 1–6, December 2011.

[152] D. S. Papailiopoulos and A. G. Dimakis. Interference alignment as a rank constrained rank minimization. In *Proc. of IEEE Global Communications Conference*, pages 1–6, 2010.

[153] S. Parkvall and D. Astely. The evolution of LTE towards IMT-Advanced. *J. Commun.*, 4(3), April 2009.

[154] S. W. Peters and R. W. Heath, Jr. Switching between antenna selection and spatial multiplexing in the nonregenerative MIMO relay channel. In *Proc. of IEEE Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, October 2008.

[155] S. W. Peters and R. W. Heath, Jr. The future of WiMAX: Multihop relaying with IEEE 802.16j. *IEEE Communications Magazine*, 47(1):104–111, January 2009.

[156] S. W. Peters and R. W. Heath, Jr. Interference alignment via alternating minimization. In *Proc. of IEEE International Conference on Acoustics Speech Signal Processing*, pages 2445–2448, Taipei, Taiwan, April 2009.

[157] S. W. Peters and R. W. Heath, Jr. Nonregenerative MIMO relaying with optimal transmit antenna selection. *IEEE Signal Processing Letters*, 15:421–424, January 2009.

[158] S. W. Peters and R. W. Heath, Jr. Cooperative algorithms for MIMO interference channels. *IEEE Transactions on Vehicular Technology*, 60(1):206–218, January 2011.

[159] S. W. Peters, A. Y. Panah, K. T. Truong, and R. W. Heath, Jr. Relay architectures for 3GPP LTE-Advanced. *EURASIP Journal on Wireless Communications and Networking*, 2009, 2009.

[160] L. P. Qian, Y. J. Zhang, and J. Huang. MAPEL: Achieving global optimality for a non-convex wireless power control problem. *IEEE Transactions on Wireless Communications*, 8(3):1553–1563, March 2009.

[161] T. Q. S. Quek, H. Shin, and M. Z. Win. Robust wireless relay networks: Slow power allocation with guaranteed QoS. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):700–713, Dec. 2007.

[162] C. Raman, G. J. Foschini, R. A. Valenzuela, R. D. Yates, and N. B. Mandayam. Half-duplex relaying in downlink cellular systems. *IEEE Transactions on Wireless Communications*, 10(5):1396–1404, May 2011.

[163] B. Rankov and A. Wittneben. Spectral efficient protocols for half-duplex fading relay channels. *IEEE Journal on Selected Areas in Communications*, 25(2):379–389, February 2007.

[164] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE Journal on Selected Areas in Communications*, 16(8):1437–1450, October 1998.

[165] M. Razaviyayn, H. Baligh, A. Callard, and Z.-Q. Luo. Joint transceiver design and user grouping in a MIMO interfering broadcast channel. In *Proc. of IEEE Conference on Information Sciences Systems*, pages 1–6, Baltimore, MD, March 2011.

[166] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo. On the degrees of freedom achievable through interference alignment in a MIMO interference channel. *to appear in IEEE Transactions on Information Theory*, 2011, http://arxiv.org/abs/1104.0992.

[167] M. Razaviyayn, M. Sanjabi, and Z. Q. Luo. Linear transceiver design for interference alignment: Complexity and computation. *to appear in IEEE Transactions on Information Theory*, 2012.

[168] S. Ren and M. Van Der Schaar. Distributed power allocation in multiuser multi-channel cellular relay networks. *IEEE Transactions on Wireless Communications*, 9(6):1952–1964, June 2010.

[169] Y. Rong and Y. Hua. Optimality of diagonalization of multi-hop MIMO relays. *IEEE Transactions on Wireless Communications*, 8(12):6068–6077, December 2009.

[170] Y. Rong, X. Tang, and Y. Hua. A unified framework for optimizing linear nonregenerative multicarrier MIMO relay communication systems. *IEEE Transactions on Signal Processing*, 57(12):4837–4851, December 2009.

[171] H. Sampath, P. Stoica, and A. J. Paulraj. Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion. *IEEE Transactions on Communications*, 49(12):2198–2206, December 2001.

[172] S. Sanayei and A. Nosratinia. Antenna selection in MIMO systems. *IEEE Communications Magazine*, 42(10):68–73, October 2004.

[173] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai. Uplink macro diversity of limited backhaul cellular network. *IEEE Transactions on Information Theory*, 55(8):3457–3478, August 2009.

[174] I. Santamaria, O. Gonzalez, R. W. Heath, Jr., and S. W. Peters. Maximum sum-rate interference alignment algorithms for MIMO channels. In *Proc. of IEEE Global Communications Conference*, pages 1–6, Miami, FL, December 2010.

[175] C. U. Saraydar, N. B. Mandayam, and D. J. Goodman. Efficient power control via pricing in wireless data networks. *IEEE Transactions on Communications*, 50(2):291–303, February 2002.

[176] A. Scaglione, P. Stoica, S. Barbarossa, G. B. Giannakis, and H. Sampath. Optimal designs for space-time linear precoders and decoders. *IEEE Transactions on Signal Processing*, 50(5):1051–1064, May 2002.

[177] D. A. Schmidt, C. Shi, R. A. Berry, M. L. Honig, and W. Utschick. Minimum mean squared error interference alignment. In *Proc. of IEEE Asilomar Conference on Signals, Systems and Computers*, pages 1106–1110, Pacific Grove, CA, November 2009.

[178] R. Schoenen, W. Zirwas, and B. H. Walke. Raising coverage and capacity using fixed relays in a realistic scenario. In *Proc. of European Wireless Conference*, pages 1–6, June 2008.

[179] D. C. Schultz and B. Walke. Fixed relays for cost efficient 4G network deployments: An evaluation. In *Proc. of IEEE International Symposium on Personal, Indoor and Mobile Radio Communcations*, Athens, Greece, September 2007.

[180] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity part 1: System description. *IEEE Transactions on Information Theory*, 51(11):1927–1938, November 2003.

[181] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity part 2: Implementation aspects and performance analysis. *IEEE Transactions on Information Theory*, 51(11):1939–1948, November 2003.

[182] H. Shen, B. Li, M. Tao, and X. Wang. MSE-based transceiver designs for the MIMO interference channel. *IEEE Transactions on Wireless Communications*, 9(11):3480–3489, November 2010.

[183] C. Shi, R. A. Berry, and M. L. Honig. Distributed interference pricing with MISO channels. In *Proc. of Allerton Conference on Communication, Control, and Computing*, pages 539–546, Monticello, IL, September 2008.

[184] C. Shi, R. A. Berry, and M. L. Honig. Local interference pricing for distributed beamforming in MIMO networks. In *Proc. of IEEE Military Communications Conference*, pages 1–6, Boston, MA, October 2009.

[185] C. Shi, R. A. Berry, and M. L. Honig. Monotonic convergence of distributed interference pricing in wireless networks. In *Proc. of IEEE International Symposium on Information Theory*, pages 1619–1623, Seoul, Korea, June 2009.

[186] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick. Distributed interference pricing for the MIMO interference channel. In *Proc. of IEEE International Conference on Communications*, pages 1–5, Dresden, Germany, June 2009.

[187] Q. Shi, M. Razaviyayn, Z. Luo, and C. He. An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel. *IEEE Transactions on Signal Processing*, 59(9):4331–4340, September 2011.

[188] Y. Shi, R. K. Mallik, and K. B. Letaief. Power control for relay-assisted wireless systems with general relaying. In *Proc. of IEEE International Conference on Communications*, pages 1–5, Cape Town, South Africa, May 2010.

[189] Y. Shi, J. Wang, K. Letaief, and R. Mallik. A game-theoretic approach for distributed power control in interference relay channels. *IEEE Transactions on Wireless Communications*, 8(6):3151–3161, June 2009.

[190] O. Simeone, O. Somekh, Y. Bar-Ness, H. V. Poor, and S. Shamai. Capacity of linear two-hop mesh networks with rate splitting, decode-and-forward relaying and cooperation. In *Proc. of Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2007.

[191] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz). Downlink macro-diversity with limited backhaul capacity. *EURASIP Journal of Wireless Communications and Networking*, 2009, 2009.

[192] O. Simeone, O. Somekh, H.V. Poor, and S. Shamai. Local base station cooperation via finite-capacity links for the uplink of linear cellular

networks. *IEEE Transactions on Information Theory*, 55(1):190–204, January 2009.

[193] S. Stanczak, M. Wiczanowski, and H. Boche. Distributed utility-based power control: Objectives and algorithms. *IEEE Transactions on Signal Processing*, 55(10):5058–5068, October 2007.

[194] S. Stanczak, M. Wiczanowski, and H. Boche. *Fundamentals of Resource Allocation in Wireless Networks*. Springer-Verlag, 2nd edition, 2009.

[195] Y. Su and M. van Der Schaar. A new perspective on multi-user power control games in interference channels. *IEEE Transactions on Wireless Communications*, 8(6):2910–2919, June 2009.

[196] C. Suh and D. Tse. Interference alignment for cellular networks. In *Proc. of Allerton Conference on Communication, Control, and Computing*, pages 1037–1044, Monticello, IL, September 2008.

[197] C. W. Sung and W. S. Wong. A noncooperative power control game for multirate CDMA data networks. *IEEE Transactions on Wireless Communications*, 2(1):186–194, January 2003.

[198] H. Sung, S.-H. Park, K.-J. Lee, and I. Lee. Linear precoder designs for K-user interference channels. *IEEE Transactions on Wireless Communications*, 9(1):291–301, January 2010.

[199] S. Talwar and A. J. Paulraj. Blind separation of synchronous co-channel digital signals using an antenna array. ii. performance analysis. *IEEE Transactions on Signal Processing*, 45(3):706–718, March 1997.

[200] X. Tang and Y. Hua. Optimal design of non-regenerative MIMO wireless relays. *IEEE Transactions on Wireless Communications*, 6(4):1398–1407, April 2007.

[201] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Transactions on Information Theory*, 45(5):1456–1467, July 1999.

[202] E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):451–460, November-December 1999.

[203] P. S. C. Thejaswi, A. Bennatan, J. Zhang, R. Calderbank, and D. Cochran. Rate-achievability strategies for two-hop interference flows. In *Proc. of Allerton Conference on Communication, Control, and Computing*, pages 1432–1439, Montecillo, IL, September 2008.

[204] M. A. Torabi and J.-F. Frigon. A decomposition approach to MIMO interference relay networks. In *Proc. of IEEE Global Communications Conference*, pages 1–6, New Orleans, LA, November-December 2008.

[205] K. T. Truong and R. W. Heath, Jr. Adaptive transmit antenna selection in MIMO amplify-and-forward relay channels. In *Proc. of IEEE*

International Conference on Acoustics Speech Signal Processing, pages 3022–3025, Dallas, TX, March 2010.

[206] K. T. Truong and R. W. Heath, Jr. Multimode antenna selection for MIMO amplify-and-forward relay systems. *IEEE Transactions on Signal Processing*, 58(11):5845–5859, November 2010.

[207] K. T. Truong and R. W. Heath, Jr. Interference alignment for multiple-antenna amplify-and-forward relay interference channel. In *Proc. of IEEE Asilomar Conference on Signals, Systems, and Computers*, pages 1–5, Pacific Grove, CA, USA, November 2011.

[208] K. T. Truong and R. W. Heath, Jr. Relay beamforming using interference pricing for the two-hop interference channel. In *Proc. of IEEE Global Communications Conference*, pages 1–5, Houston, TX, December 2011.

[209] K. T. Truong and R. W. Heath, Jr. Cooperative transmit precoding for the relay interference broadcast channel. *in preparation for submission to IEEE Transactions on Vehicular Technology*, February 2012.

[210] K. T. Truong, P. Sartori, and R. W. Heath, Jr. Cooperative algorithms for MIMO amplify-and-forward relay networks. *submitted to IEEE Transactions on Signal Processing*, December 2011, Available: http://arxiv.org/abs/1112.4553.

[211] D. Tse and P. Visnawath. *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[212] M. Vajapeyam and U. Mitra. A hybrid space-time coding scheme for cooperative networks. In *Proc. of Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2004.

[213] E. C. Van der Meulen. Three-terminal communication channel. *Advances in Appliced Probability*, 3:120–154, 1971.

[214] H. Viswanathan and S. Mukherjee. Performance of cellular networks with relays and centralized scheduling. *IEEE Transactions on Wireless Communications*, 4(5):2318–2328, September 2005.

[215] B. Wang, J. Zhang, and A. Host-Madsen. On the capacity of MIMO relay channels. *IEEE Transactions on Information Theory*, 51(1):29–43, January 2005.

[216] R. Wang and G. B. Giannakis. Approaching MIMO channel capacity with soft detection based on hard sphere decoding. *IEEE Transactions on Communications*, 54(4):587–590, April 2006.

[217] T. Wirth, V. Venkatkumar, T. Haustein, E. Schulz, and R. Halfmann. LTE-Advanced relaying for outdoor range extension. In *Proc. of IEEE Vehicular Technology Conference*, pages 1 –4, September 2009.

[218] L. Xiao and L. Cuthbert. Multi-cell non-cooperative power allocation game in relay based OFDMA systems. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, Barcelona, Spain, April 2009.

[219] C. Xing, S. Ma, and Y.-C. Wu. Robust joint design of linear relay precoder and destination equalizer for dual-hop amplify-and-forward MIMO relay systems. *IEEE Transactions on Signal Processing*, 58(4):2273–2283, April 2010.

[220] Jiaming Xu, Jun Zhang, and J. G. Andrews. On the accuracy of the Wyner model in cellular networks. *IEEE Transactions on Wireless Communications*, 10(9):3098–3109, September 2011.

[221] S. Xu and Y. Hua. Source-relay optimization for a two-way MIMO relay system. In *Proc. of IEEE International Conference on Acoustics Speech Signal Processing*, pages 3038–3041, Dallas, TX, March 2010.

[222] W. Xu, X. Dong, and W.-S. Lu. MIMO relaying broadcast channels with linear precoding and quantized channel state information feedback. *IEEE Transactions on Signal Processing*, 58(10):5233–5245, October 2010.

[223] Y. Yang, H. Hu, J. Xu, and G. Mao. Relay technologies for WiMax and LTE-Advanced mobile systems. *IEEE Communications Magazine*, 47(10):100–105, October 2009.

[224] R. D. Yates. A framework for uplink power control in cellular radio systems. *IEEE Journal on Selected Areas in Communications*, 13(7):1341–1347, September 1995.

[225] C. M. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran. On feasibility of interference alignment in MIMO interference networks. *IEEE Transactions on Signal Processing*, 58(9):4771–4782, September 2010.

[226] E. Yilmaz, R. Knopp, and D. Gesbert. On the gains of fixed relays in cellular networks with intercell interference. In *Proc. of IEEE Workshop on Signal Processing Advances in Wireless Communications*, pages 603–607, June 2009.

[227] W. Yu and T. Lan. Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Transactions on Signal Processing*, 55(6):2646–2660, June 2007.

[228] R. Zakhour, Z. K. M. Ho, and D. Gesbert. Distributed beamforming coordination in multicell MIMO channels. In *Proc. of IEEE Vehicular Technology Conference*, pages 1–5, Barcelona, Spain, April 2009.

[229] H. Zhang and H. Dai. Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks. *EURASIP Journal of Wireless Communications and Networking*, 2004(2):222–235, 2004.

[230] J. Zhang and J. G. Andrews. Adaptive spatial intercell interference cancellation in multicell wireless networks. *IEEE Journal on Selected Areas in Communications*, 28(9):1455–1468, December 2010.

[231] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr. Networked MIMO with clustered linear precoding. *IEEE Transactions on Wireless Communications*, 8(4):1910–1921, April 2009.

[232] R. Zhang, C. C. Chai, and Y.-C. Liang. Joint beamforming and power control for multiantenna relay broadcast channel with QoS constraints. *IEEE Transactions on Signal Processing*, 57(2):726–737, February 2009.

[233] R. Zhang, Y.-C. Liang, R. Narasimhan, and J. M. Cioffi. Approaching MIMO-OFDM capacity with per-antenna power and rate feedback. *IEEE Journal on Selected Areas in Communications*, 25(7):1284–1297, September 2007.

[234] J. Zhao, I. Hammerstroem, M. Kuhn, A. Wittneben, M. Herdin, and G. Bauch. Coverage analysis for cellular systems with multiple antennas using decode-and-forward relays. In *Proc. of IEEE Vehicular Technology Conference*, pages 944–948, April 2007.

[235] S. Zhou, H. Xiao, and Z. Niu. Distributed power control for interference-limited cooperative relay networks. In *Proc. of IEEE International Conference on Communications*, pages 1–6, Dresden, Germany, June 2009.

Vita

Kien T. Truong was born and raised in Hanoi, Vietnam. He received B.S. degree in Electronics and Telecommunications from Hanoi University of Technology, Hanoi, Vietnam, in 2002 and the M.Sc. degree in Electrical Engineering from The University of Texas at Austin, Austin, TX, USA, in 2008. From 2002 to 2006, he was a research engineer in the Department of Wireless Communications in the Research Institute of Posts and Telecommunications, Hanoi, Vietnam. He is a recipient of the Vietnam Education Foundation Fellowship in 2006-2011. He is also a recipient of the David Bruton Jr. Fellowship in 2007-2008. His research interests include link adaptation and interference management for relay-aided communication systems and capacity analysis of wireless ad hoc networking.

Email address: kientruong@utexas.edu

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.